# On a Fokker–Planck equation coupled with a constraint

Analysis of a lithium-ion battery model

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Dr. Rer. Nat.
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät II
Humboldt-Universität zu Berlin

von
**Dipl.-Math. Robert Huth**

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II:
Prof. Dr. Elmar Kulke

Gutachter:
1. Prof. Dr. Alexander Mielke, HU-Berlin
2. Prof. Dr. Wolfgang Dreyer, TU-Berlin
3. Prof. Dr. Robert Denk, Uni-Konstanz

**eingereicht am:** 11.01.2012
**Tag der mündlichen Prüfung:** 31.05.2012

## Danksagung

Ich möchte mich hier ganz herzlich bei den Menschen bedanken, die mich bei dieser Arbeit begleitet und unterstützt haben. Zuerst müssen hier meine beiden Betreuer Prof. Dr. Wolfgang Dreyer und Prof. Dr. Alexander Mielke erwähnt werden. Beide haben mich in der Erstellung dieser Arbeit unterstützt. Der Eine von der physikalischen Seite, der Andere von der mathematischen Seite. Von beiden habe ich viel gelernt und bin für deren Beistand sehr dankbar.

Ich möchte mich auch bei Dr. Joachim Rehberg bedanken. Er stand mir beim Erlernen von mathematischen Methoden um die sektoriellen Operatoren, wie ein Mentor zur Seite. An diesen Techniken habe ich viel Spaß gefunden und ich habe die Arbeit mit ihm genossen. Des Weiteren bedanke ich mich bei meinen Freunden und Kollegen Dr. Rüdiger Müller, Margarita Naldzhieva, Anita Schirrmacher und Gerhard Grübler, die in verschiedenen Phasen dieses Manuskript Korrektur gelesen haben. Ihre Anmerkungen waren für mich sehr wertvoll. Neben den oben schon genannten, bedanke ich mich auch bei allen anderen Kollegen des Weierstraß-Instituts. Namentlich möchte ich erwähnen: Clemens Guhlke, Dr. Dirk Peschka, Ernst Höschele, Dr. Gonca Aki, Dr. Frank Duderstadt, Thomas Petzold, Lukas Wilhelm, Dr. Sebastian Heinz, Sebastian Jachalski, Lars Giere, Christoph Grützmacher, Dr. Alexander Linke, Volker Schloßhauer und Ina Hohn. Diese Kollegen haben mein tägliches Arbeitsklima positiv beeinflusst. Letztendlich möchte ich mich recht herzlich bei meiner Familie und meiner Freundin für die Kraft und Motivation bedanken, die sie mir währenddessen geschenkt haben.

**Abstract**

We discuss two models which describe the charging and discharging of a lithium-ion battery and especially the hysteretical behaviour therein. We give an overview on the modelling process for a *discrete many particle model* and a *continuous many particle model*. The former results in an axiomatic description of macroscopic quantities while the latter gives a nonlinear Fokker–Planck equation. We analyse the *discrete many particle model* and later compare its properties to the *continuous many particle model*.

The nonlinear Fokker–Planck equation is analysed with respect to existence and uniqueness of solutions as well as qualitative behaviour of solutions. The nonlinearity in this partial differential equation stems from a coefficient which depends on the solution first non-local and second in a higher order, i.e. it depends on the solution as a function in $\mathrm{C}(\bar{\Omega})$ and not only in $\mathrm{L}^2(\Omega)$. We use interpolation spaces and semigroups generated from sectorial operators to show the existence and uniqueness of solutions locally in time. We also show the positivity and regularity of solutions. The global existence in time relies on estimates for the dissipation of an energy. The suitable energy is related to the $L \log L$ norm and so a Gagliardo-Nirenberg inequality is needed to connect this back to $\mathrm{L}^2(\Omega)$ estimates. It turns out that the conditions for global in time existence of solutions are exactly what one expects from a physical point of view. The condition is that the loading state of the battery shall never exceed the state of being totally empty or totally full.

The investigation of the qualitative behaviour of solutions to the nonlinear Fokker–Planck equation is done by numerical simulations and matched formal asymptotics. In order to verify the numerical experiments, we show convergence of the numerical solutions. For this we exploit that the discrete operators inherit properties from their original counterparts. We do not rely on high regularity of the solution or of the initial data.

We document the behaviour of solutions to concentrate at certain points in numerical experiments. Together with the results from formal asymptotic expansions, we see that the limiting behaviour for a certain scaling of the appearing parameters results in the formation of Dirac measures. This is congruent with known results for the linear Fokker–Planck equation. In the nonlinear case at hand we observe next to the mentioned concentration of solutions a hysteresis in the dynamics of global quantities. The evolution of the global quantities, which we observe in numerical simulations, is the same that results from the *discrete many particle model* and one observes hysteretic behaviour in macroscopic quantities.

## Zusammenfassung

In dieser Arbeit untersuchen wir zwei Modelle, die das Laden und Entladen einer Lithium-Ionen Batterie beschreiben. Beide Modelle zielen darauf ab eine Hysterese in dem Spannungs-Ladungs-Verlauf widerzuspiegeln. Wir skizzieren den Modellierungsprozess von einem *diskreten vielteilchen Modell* und einem *kontinuierlichen vielteilchen Modell*. Das erste führt zu einer axiomatischen Beschreibung der Evolution makroskopischer Größen, während das zweite in eine nichtlineare Fokker–Planck Gleichung mündet. Das *diskrete vielteilchen Modell* wird analysiert um das *kontinuierliche vielteilchen Modell* zu motivieren und es später mit diesem vergleichen zu können.

Wir untersuchen die Existenz und Eindeutigkeit von Lösungen einer nichtlinearen Fokker–Planck Gleichung, sowie deren qualitative Eigenschaften. Die Nichtlinearität dieser partiellen Differentialgleichung kommt von einem Koeffizienten, der von der Lösung auf eine nichtlokale Weise abhängt. Dieser Koeffizient hängt von der Lösung als Funktion in $\mathrm{C}(\bar{\Omega})$ ab. Da wir in den Räumen $\mathrm{H}^1(\Omega)$ und $\mathrm{H}^{-1}(\Omega)$ arbeiten, brauchen wir Interpolationsräume und Halbgruppen sectorieller Operatoren um einen semilinearen Charakter der partiellen Differentialgleichung zu erhalten und auszunutzen. Somit zeigen wir letztendlich Eindeutigkeit und Existenz von Lösungen in der Zeit. Wir zeigen auch die positivität von Lösungen und deren Regularität. Um globale Existenz zu erhalten, müssen wir die Dissipation einer mit dem Modell verknüpften Energie abschätzen. Diese Energie ist verwandt mit der $L \log L$ Norm. Um also Beschränkheit in $\mathrm{L}^2(\Omega)$ zu erhalten, müssen wir eine Gagliardo-Nirenberg Ungleichung benutzen. Es stellt sich heraus, dass die notwendigen und hinreichenden Bedingungen zur globalen Existenz von Lösungen solche sind, welche man aus einer physikalischen Sicht erwarten würde. Man benötigt den Umstand, dass der Ladezustand der Batterie niemals die Werte *voll* oder *leer* über- oder unterschreitet.

In einer Reihe von numerischen Experimenten untersuchen wir das qualitative Verhalten von Lösungen. Wir zeigen die Konvergenz der numerischen Lösungen und verifizieren so unsere numerischen Experimente. Dabei nutzen wir aus, dass die diskretisierten Operatoren ihre Eigenschaften von deren jeweiligen Gegenstücken im ursprünglichen Problem erben. Wir nutzen letztendlich ähnliche Techniken wie bei der lokalen Existenz von Lösungen der nichtlinearen Fokker–Planck Gleichung. Wir brauchen dafür keine starke Regularität der exakten Lösung oder des Anfangswertes. Dafür ist die Konvergenz der numerischen Lösungen, die wir bekommen, von nur abstrakter Natur und wir können keine Konvergenzraten angeben.

Wir zeigen in numerischen Experimenten die Tendenz von Lösungen sich um bestimmte Punkte zu konzentrieren. Wir zeigen auch den Einfluß der Parameter auf dieses Verhalten. Zusammen mit Ergebnissen aus der formalen Asymptotik zeigt dies für eine bestimmte Wahl von Parameter-Skalierungen, dass Lösungen gegen Dirac-Maße konvergieren. In diesem Grenzverhalten scheint das System wieder durch die Evolution von makroskopischen Größen beschrieben zu werden, welche wir in dem *diskreten vielteilchen Modell* beschrieben haben und man kann in diesen makroskopischen Größen eine Hysterese beobachten.

# Contents

# 1 Introduction

The work in your hands revolves around equations which arise in the modelling of a rechargeable lithium-ion battery. As reversible storage devices, lithium-ion batteries are in the focus of modern technological development. The search for a replacement of fuel as a mobile energy storage in cars as well as the wide range of mobile electronic devices like cell phones, portable computers and many more demand for an energy storage which is small and light, i.e. has high energy density. This list of desired properties can be prolonged by robustness, affordability, safety and fast rechargeability. We certainly do not want to wait hours at a highway gas (electrical power) station while recharging our car's battery.

What makes lithium so attractive for storage and reproduction of electrical energy is that its electrochemical potential is the lowest among all elements in the periodic table, which permits the highest theoretical voltage. Additionally it has after hydrogen and helium the lowest atomic mass, which results in extremely high power-to-weight ratios for this type of batteries. In practice lithium-ion batteries even show no memory effect and a comparable low loss of charge when not used. Another property of lithium-ion batteries is the occurrence of a voltage plateau in the voltage-capacity behaviour while loading and unloading. A typical experiment reflecting these plateaus is shown in Figure 1.1. There we see a hysteretic behaviour, because the path of the curve in the voltage-capacity plot is different for charging and discharging. The occurrence of a voltage plateau is a desirable property, because it results to near constant voltages over a wide range of the capacity which is needed by most modern electronic devices. The vertical distance of the two plateaus on the other hand implies a loss of energy. When discharging less energy can be retrieved than the amount that had to be applied while charging.

In order to optimise the properties of a battery, a deep understanding of the underlying physical and chemical processes is needed. We need models that reflect the charging and discharging behaviour and give rise to how this behaviour can be influenced by the design of the battery. In this work we are interested in the hysteresis as observed in Figure 1.1. The understanding of this phenomenon is still very young. The probably earliest work on lithium-ion batteries with an iron-phosphate cathode goes back to Padhi, Nanjundaswamy and Goodenough [32] in 1997. We discuss two different models aimed at the description of the charge-discharge hysteresis and then highlight their advantages and disadvantages. Following that, we introduce and analyse a third model, stated by Dreyer, Guhlke and Herrmann in [11], which results in a partial differential equation (PDE) of Fokker–Planck type. As a first sanity check we answer the non-trivial questions of existence and uniqueness of solutions positively. A PDE admitting no solution could clearly not reflect any real world behaviour. On the other hand the occurrence

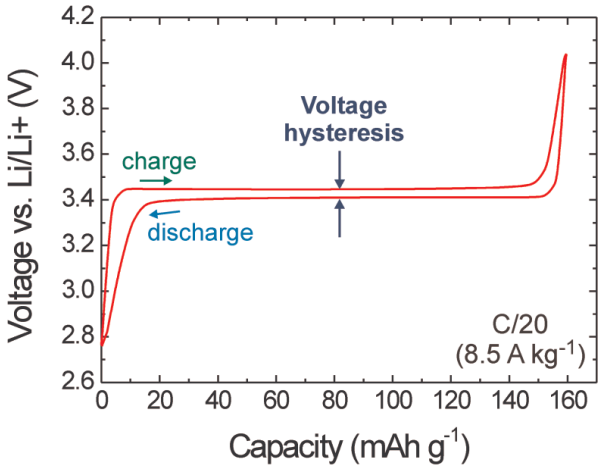



Figure 1.1: Measured voltage capacity behaviour in experiments for a lithium-ion battery, note the voltage plateaus at different heights while charging and discharging, taken from [13].

of multiple solutions would indicate that a selective mechanism is missing in the model because (roughly speaking) there is only one reality in nature. We further investigate the qualitative behaviour of the solutions with mathematical tools and the way it is influenced by the appearing parameters.

The models we discuss in this work stem from a modelling of the underlying physical and chemical processes. This is in contrast to models whose purpose is to only reflect the evolution of the capacity and voltage while (dis)charging, see for example [37]. These can be useful to predict the usage of a battery in a satellite before deployed to space or to deduce the capacity of a battery from measurements. In either case we would not need the connection to the exact inner physical and chemical processes in the battery. Our objectives however, are different.

It is common believe that the voltage plateaus while (dis)charging have their origin in a phase transition, taking place in the cathode (positive electrode). The cathode typically consists of nano-sized particles of metal oxide. While discharging, lithium atoms are intercalated in these particles. Further detailed description will be given later on. The earliest attempt to describe the almost horizontal voltage plateaus describes a phase separation inside these particles, see Padhi et al. [32]. However, they can predict the hysteretic behaviour only unsatisfactorily. A new work by Dreyer, Jamnik, Guhlke, Huth, Moskon and Gaberscek [14] reveals that the origin of the voltage plateaus is not the phase separation inside a single particles. It rather stems from the interaction of many nanosized particles in the cathode and their changing chemical properties while the lithium is intercalated in them. We give an overview of the modelling process of a *discrete many particle model* and deduce resulting properties. The detailed modelling process can be found in [14]. Additionally we compare the *discrete many particle model*

to a class of models which describe the phase change in one single particle.

The battery model which is in the focus of this work is a *continuous many particle model* and has its origin in the *discrete many particle model*. The *continuous many particle model* is derived in detail by Dreyer, Guhlke and Herrmann in [11]. It models the large number of interacting nanosized particles with a probability density over the possible loading states. The probability density allows to consider an ensemble of a large number of particles which are in practice more than $10^{10}$ and it also allows to easily include entropic effects. The actual state of any specific particle in the cathode is not captured any more, but rather the interactions in the ensemble. The resulting model is a nonlinear PDE of Fokker–Planck type.

The general Fokker–Planck equation in one dimension is a drift-diffusion equation of the form

$$\partial_t u(x,t) = \partial_x\Big(\frac{1}{\tau}\ u(x,t)\partial_x H(x)\Big) + \nu^2 \partial_{xx} u(t,x)\ ,$$

where $x \in \Omega \subset \mathbb{R}$, $\tau, \nu > 0$ and $H$ is a potential describing an energy landscape of particles with varying property $x \in \Omega$. We neglect possible boundary conditions for now. This equation goes back to Kolmogoroff [24] and is also called the *Kolmogoroff forward equation*. It models stochastic processes like chemical reaction or particles under Brownian motion as formulated by Kramers [17]. Due to its linear parabolicity, existence and uniqueness of solutions are easily affirmed. In some cases depending on $H$ there are even explicit solutions available, see Englefield [15], where a nice transformation into an equation of Schrödinger-type is used. Furthermore reaction rates and the convergence to Dirac measures for asymptotic regimes of $\nu$ and $\tau$ have been derived formally by Kramers [17]. Recently, in 2010, Peletier Savaré and Veneroni in [34] have derived the reaction rates and the mentioned convergence even rigorously via Gamma-convergence techniques. An interesting formulation of the Fokker–Planck equation as a Wasserstein gradient flow was introduced by Jordan, Kinderlehrer and Otto [22] in 1998. Anrich, Mielke, Peletier, Saveré and Veneroni in [1] have reproven convergence to Dirac measures and reaction rates in 2011 exploiting the Wasserstein gradient-flow structure. The employed techniques in both cases exploit the existence of a known unique stationary state, which are equivalently minimisers of some energy functional.

In the variant of the Fokker–Planck equation which we deal with, the potential $H$ depends not only on the spatial variable, but also on time and the solution itself. The difficulty arises from a coefficient $\mathbb{L}$ in $H$ of the form

$$\mathbb{L}(u(t)) = \int_\Omega u(t,x)\mu(x)\,\mathrm{d}x\ +\ \nu^2\big(\ u(t,1) - u(t,0)\ \big)\ ,$$

where the domain is $\Omega = ]0,1[$, and $\mu$ is a function which we will specify later on. This coefficient in $H$ depends nonlocally on $u$ as a continuous function and turns the linear Fokker–Planck equation into a nonlinear one. At first sight, the Laplacian can be considered as a mapping from $\mathrm{H}^2(\Omega)$ functions into $\mathrm{L}^2(\Omega)$ functions and the semilinear term, $N(v) := \partial_x\Big(\beta\ v(x)\partial_x H(\mathbb{L}(v),x)\Big)$, maps $\mathrm{H}^1(\Omega)$ functions into $\mathrm{L}^2(\Omega)$ functions

(we omit a possible dependence on time for now). Consequently the nonlinearity is of lower order than the Laplacian and hence the equation is a semilinear parabolic partial differential equation.

It turns out that the boundary conditions, which arise from the model, force us to look at this PDE in the pair of spaces $(\mathrm{H}^1(\Omega),\mathrm{H}^{-1}(\Omega))$ and the Laplacian as a map from one into the other. Hence we have to reconsider the nonlinear term $N$ and how it maps into $\mathrm{H}^{-1}(\Omega)$. In this consideration the nonlinear term $N$ maps from $\mathrm{C}(\bar{\Omega})$ into $\mathrm{H}^{-1}(\Omega)$. In order to work on Sobolev spaces and still exploit the good parabolic properties of the Laplacian and the semilinear character of the PDE we have to use interpolation spaces as described in the textbook by Lunardi [28]. The naive usage of the embedding $\mathrm{H}^1(\Omega) \hookrightarrow \mathrm{C}(\bar{\Omega})$ and hence the interpretation $N : \mathrm{H}^1(\Omega) \to \mathrm{H}^{-1}(\Omega)$ would destroy the semilinear character and result in a fully nonlinear equation. Complex interpolation spaces permit noninteger Sobolev exponents and so we can use the embedding $\mathrm{H}^{\frac{1}{2}+\varepsilon} \hookrightarrow \mathrm{C}(\bar{\Omega})$, which holds in one dimension and hence $N : \mathrm{H}^{\frac{1}{2}+\varepsilon} \to \mathrm{H}^{-1}(\Omega)$. The space $\mathrm{H}^{\frac{1}{2}+\varepsilon}$ is of lower regularity than $\mathrm{H}^1(\Omega)$ which is the domain of the Laplacian and so we are able to use techniques for semilinear equations. Semigroup techniques, see [28], provide existence and uniqueness of solutions local in time. The global in time existence of solutions is a more delicate task and rises the need to employ a more complicated local in time existence result.

The needed assumptions on the given data for solutions to exist for all times turn out to match with what one expects from the real world interpretation of the model. The overall loading state must not exceed the physical boundaries *full* or *empty*. The deduction of it relies on a priori estimates for the dissipation $\mathcal{D}$ which is the negative change of a suitable energy $\mathcal{A}$ along solution trajectories $u : [0,T] \to \mathrm{H}^1(\Omega)$. The energy $\mathcal{A}$ is the same as the one which appears in gradient-flow structure for the original Fokker–Planck equation and also showed up in the modelling process. In our nonlinear version we cannot guarantee that this energy decreases at all times and so $\mathcal{D}$ must be estimated as good as possible from above in order to guarantee that $\mathcal{A}$ stays a priori bounded global in time. The boundedness of $\mathcal{A}$ then helps us to deduce the boundedness of solutions $u$. The delicate part is that solutions live in spaces $\mathrm{H}^1(\Omega)$ or at least in $\mathrm{H}^{\frac{1}{2}+\varepsilon}$, while the resulting estimates for $\mathcal{A}$ give only boundedness of $u$ in $\mathrm{L}^2$. This creates the need to modify standard semilinear existence results for initial data of low regularity such that the initial value is in $\mathrm{L}^2(\Omega)$ but the solution is for positive times in $\mathrm{H}^1(\Omega)$.

We also show convergence of finite element solutions. For this we exploit that the discretised operators inherit the properties of their original versions. This strategy is similar to work done by Bakaev in [2] and Lubich and Osterman in [27]. However in their work semigroup techniques are used in order to provide numerical error estimates for PDE's with initial values of low regularity, when on the other hand we use these techniques to pay for the low regularity of solutions, i.e. $\mathrm{H}^1(\Omega)$ but not $\mathrm{H}^2(\Omega)$

Solutions to the nonlinear version of the Fokker–Planck equation conserve their mass and positiveness and can thus be interpreted as probability measures for all times. We document in numerical experiments qualitative properties like the convergence to Dirac measures and other limiting behaviour for a certain choice of limiting parameters. To-

gether with results from formal matched asymptotics we see that solutions tend to concentrate mass around two points in the property space $\Omega$ and we can deduce formally their shape and evolution, which is in the limit similar to the effective model which we deduced from the *discrete many particle model*. The rigorous derivation of these properties in [1] and [34] for the linear Fokker–Planck equation were not transferable to our problem as we lack the existence of a unique stationary state. It can be seen easily that a whole family of these stationary states exists. Additionally we cannot use typical transformations of our problem, because the nonlinear term gets even more complicated and destroys the gain which arises in the linear formulation. In a very recent work by Herrmann, Niethammer and Velázquez [19] even more limiting behaviours for various scales of parameters were formally derived. This shows that this one-dimensional nonlinear partial differential equation is still a source for interesting questions as the limit behaviour is still not fully understood.

The following work is divided as follows: We start in Chapter 2 with the basic modelling of a lithium ion battery. We give a short overview of an assembly of a battery and discuss two models for (dis)charging, the *core shell* model and the *discrete many particle model*. While we only sketch the former, we analyse the latter for qualitative properties. The properties derived in Chapter 2 will be later compared to the *continuous many particle model*.

In Chapter 3 we give an overview on the model given by Dreyer, Guhlke and Herrmann [11] and we show their derivation briefly. We leave a lot of physical details aside. The close connection to the *discrete many particle model* gets clear right from the beginning. We reflect from [11] that once material properties are derived from basic physical principles, a nonlinear partial differential equation of Fokker–Planck type arises. Additionally first elementary properties are mentioned here.

Chapter 4 is devoted to analytical investigations of the nonlinear Fokker–Planck equation as stated in Chapter 3. It presents a joint work of Dreyer, Huth, Mielke, Rehberg and Winkler in [13]. Existence and uniqueness is shown and the derivation of necessary and sufficient blow-up criteria is given. We also reveal that solutions are positive and in $\mathrm{H}^1(\Omega)$ for all positive times and we characterise the family of stationary solutions. The analysis of blow-up criteria exploits that the problem is highly connected to a special energy $\mathcal{A}$ that showed up in the modelling process. Similar techniques that allow to derive blow up criteria give rise to the convergence to stationary states at infinite time for suitable data.

We investigate the qualitative behaviour by numerical experiments in Chapter 5. At the beginning we explain the used numerical method. Then convergence of numerical solutions to the exact solution is shown. This is done in a general way with the aim to be applicable to other semilinear parabolic problems. For the convergence result we employ similar techniques as we used to show the short time existence. We carry over properties of the operators in the continuous PDE to their discretised versions and derive a priori estimates and convergence of numerical solutions. Then we carefully investigate the convergence to Dirac measures and the formation of hysteretic behaviour under the influence of changing parameters with numerical experiments.

The interesting behaviour of solutions for a specific limit of parameters is further

investigated in Chapter 6 with the method of formal matched asymptotics. The derived results are congruent with those of numerical experiments. It reflects the behaviour of the solutions to concentrate mass around certain points such that the limit is a Dirac measure.

The thesis finishes with a conclusion which summarises the results in a nutshell in Chapter 7. We also highlight questions which either are still open or are induced by the findings presented in this work. The Conclusion is followed by an Appendix where we list frequently used estimates. We do so for easier referenciation and self containment.

# 2 Earlier Models

In this chapter we will give an overview of two battery models. In the order they are presented here, these two models show an evolution of a modelling process. The main model of this work, which is found in Chapter 3, is a result of this evolution.

## 2.1 The Common Idea of Presented Models

The models discussed in this chapter as well as the model in Chapter 3 describe the same type of a Li-ion battery. Thus all share a common base which we present here.

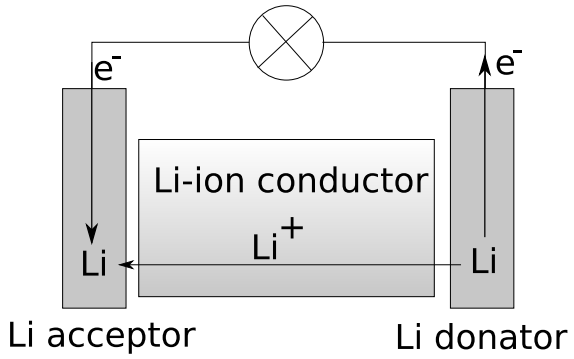



Figure 2.1: Li-ion battery scheme for discharging

As sketched in Figure 2.1 a Li-ion battery consists of two electrodes which are separated by a Li-ion conducting material. Electrodes are usually called anode and cathode depending on their role as electron donator or electron acceptor. Since this role can change in a rechargeable battery, we do not use this naming here.

Both electrodes can store lithium atoms. The materials for those two electrodes are chosen in such a way, that it is energetically more favourable for the lithium atoms to be in one electrode compared to the other. We call the electrode where it is less favourable for lithium atoms to be, lithium donator, and the other lithium acceptor. This naming is motivated from their roles when discharging the battery. Note that this role does also change when charging, but we avoid using the more strictly defined words *cathode* and *anode*. The only way lithium ions can travel from one electrode to the other is by passing through the Li-ion conductor. The crucial point is that this is a conductor for lithium ions, but not for lithium atoms nor for electrons. Thus the lithium atoms must each emit an electron on the Li-donator side in order to become ions and be able to

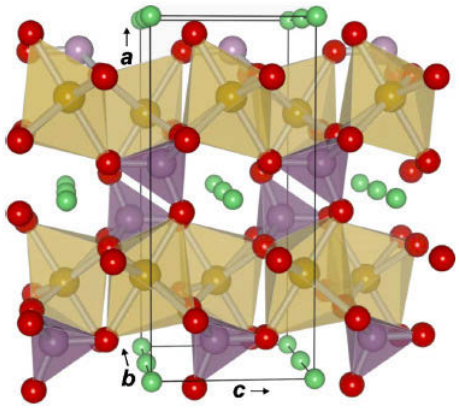



Figure 2.2: Scheme of $FePO_4$ lattice with free interstitial lattice sites, green, which can be occupied by lithium atoms, taken from [29].

pass through the electrolyte. While the ions travel through the electrolyte, the electrons then travel through an outer circuit and can be used as electric energy. The ion and the electron recombine on the Li-acceptor side to become a lithium atom. Finally this lithium atom is then stored in the Li-acceptor electrode.

The just described process is the process of discharging. While charging the process takes place in reversed order. To enforce the reverse process one has to apply an electric tension to the electrodes, charging the Li-donator electrode negative compared to the acceptor electrode. When the applied voltage is high enough, the tendency of the system to neutralise this charge is higher than the energetical benefit for lithium atoms to be in the lithium acceptor electrode and the battery gets charged.

The Li-donator electrode of the battery we want to model consists of pure lithium. The acceptor electrode is made of iron-phosphate, i.e. $FePO_4$. This material can store lithium atoms in interstitial lattice sites, as depicted in Figure 2.2. To be more precise, we consider the Li-acceptor electrode to consist of many nanosized $FePO_4$ particles, which are packed on a current collector as sketched in Figure 2.3. All these particles are connected in an electrically conducting way with the current collector and with each other. The current collector usually consists of a highly (electrical) conducting material such as aluminium foil.

It is well known that the electrical conductivity of $FePO_4$ is very poor. Hence one adds carbon black to the ensemble of ironphosphate particles. This is ideally done by coating the particles as described in the patent [36]. The construction of the $FePO_4$ electrode considered in this work is in consensus with those considered in the chemical community, see for example Newman and Srinivasan [38] or Delmas et. al [8].

The state of charge of the battery can be measured by

$$\ell = N_{\mathrm{Li}}/N_{\mathrm{V}}. \tag{2.1}$$

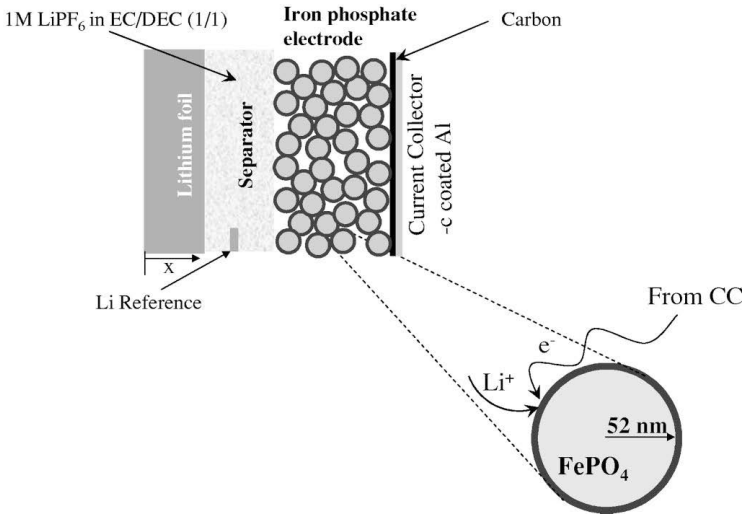



Figure 2.3: Acceptor electrode consisting of an ensemble of nanosized $FePO_4$ particles, taken from [38].

Here $N_V$ is the total number of interstitial lattice sites in the lithium acceptor electrode, which can possibly hold lithium atoms. The total number of intercalated lithium atoms therein is $N_{Li}$. Thus $0 \leq \ell \leq 1$ and $\ell = 0$ means that the battery is fully charged and $\ell = 1$ means fully discharged.

The voltage of the battery is related to the concentration of lithium in the lithium acceptor electrode. More precisely, it is related to the lithium concentration on the surface of the iron phosphate electrode exposed to the electrolyte, because here lithium atoms can possibly enter or leave the electrode. A difference in chemical properties to lithium atoms, stored in the surface of the other electrode, is the source of the electric tension or voltage of the battery.

We assume that it suffices to describe the (de)intercalation process of lithium into the acceptor electrode in order to model the (dis)charging behaviour of the battery. The process of lithium intercalation is shortly called lithiation. In order to simplify the model we assume the lithium donator electrode to be made of pure lithium. The potential of the lithium donator electrode then stays the same at any state of charge. Of course extensions to model a complexer lithium donator electrode are possible. The fast diffusion of lithium ions through the electrolyte compared to the diffusion speed of lithium atoms in the $FePO_4$ lattice allows us to consider only the lithiation of the $FePO_4$ electrode.

The material $FePO_4$ undergoes a phase change while lithiation, as observed in [32]. It is believed that this is the source of the interesting charge and discharge behaviour mentioned in the introduction. All the models presented in this work are an attempt to find the way in which this phenomenon influences the charge-discharge characteristics of a battery.

## 2.2 Core Shell Model

The core shell models are widely used in the chemical community to explain the loading behaviour of batteries. They are able to explain the occurrence of voltage plateaus and hysteresis in the voltage capacity plots while charging and discharging. This section explains the principles of core shell models and in the end will discuss its shortcomings. Models derived with this idea can still differ a lot. For further reference we refer to Newman and Srinivasan [39, 38] or Dreyer et al. [9]. A recent work with a lot of references in the context of core shell models is a paper by Delacourt and Safari [37].

A core shell model incorporates the idea, that two phases of lithium iron phosphate manifest in one nanosized iron phosphate particle. One Phase has a high and the other phase a low lithium concentration. The tendency of iron phosphate to develop those two phases during lithiation was described in [32].

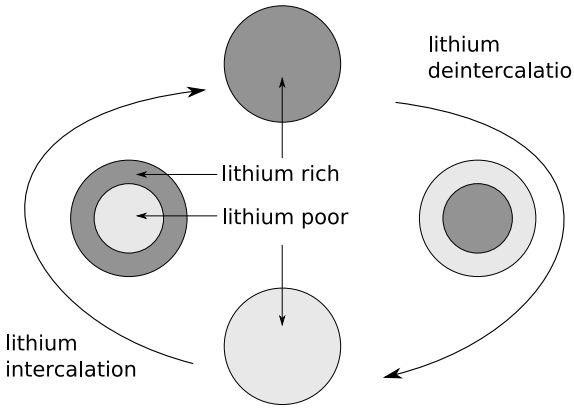



Figure 2.4: scheme of lithium intercalation and deintercalation according to the core shell model.

When lithium enters the particle through the outer surface, it is assumed that a lithium rich phase grows from the outer surface into the inside. Thus, as depicted in Figure 2.4, the particle consists of an outer shell which is lithium rich and an inner core which is lithium poor. For simplicity one assumes that the particle and the core have the shape of a ball.

The total filling degree of a particle $\ell$, see (2.1), increases while the shell of lithium rich iron phosphate grows and the lithium poor core shrinks. At the same time the lithium concentration at the surface, which is responsible for the potential or voltage of this particle, stays roughly spoken the same. The crucial point is that one assumes that all particles behave the same. Thus the state of charge of the whole battery is described by the filling degree of a single reference particle and its potential gives rise to the voltage of the whole battery. Consequently this model is able to describe the occurrence of voltage plateaus in Figure 1.1. The process is not symmetric. During lithiation we have a growing shell with high lithium concentration and a lithium poor core while during

delithiation we have a growing shell with low lithium concentration and a lithium rich core. This explains the hysteretic effect in the voltage-capacity characteristic of the battery, seen in Figure 1.1.

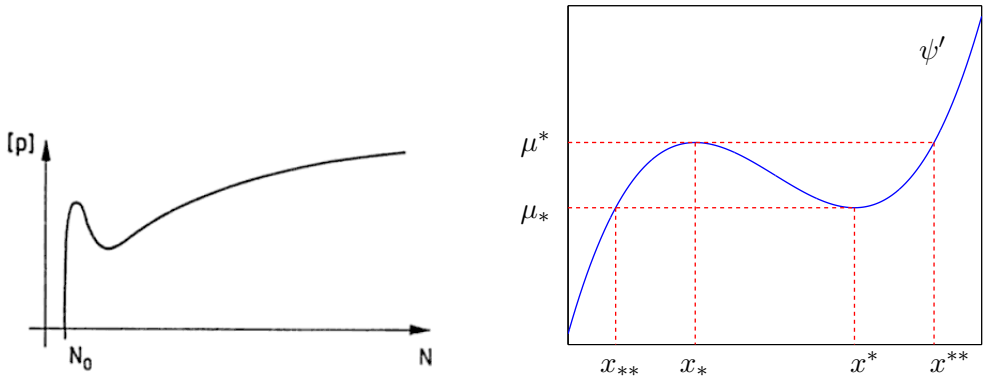
So why do we need other models? As described by Delmas et al. [8] and the references therein, recent advances have shown that the explained core-shell behaviour is not likely to occur. In fact they are not able to find the occurrence of nanosized particles having such a core shell structure in experiments. According to Delmas et al., the $FePO_4$ lattice has a special three-dimensional structure. In this structure lithium can only move in one preferred direction, depending on the grid orientation. One speaks of so called tubes in the lattice, in which the lithium atoms can move. It is believed, that lithium atoms thus have a preferred drift direction in one whole nanosized particle, which makes the formation of a shell unlikely. In the same work it is said that the experimental charge/discharge behaviour can not be predicted satisfactorily by core shell models.

In a work by Dreyer, Gabercek, Guhlke, Huth and J. Jamnik [9] the behaviour of a core shell particle is deduced from basic material properties and physical principles. The authors show that for small particle sizes no core shell configuration would be stable any more. This is due to high surface tension on the interface between the inner core and the outer shell. This means that reducing the particle size in a battery would change the charging characteristic significantly, which contradicts the experiments found in [38]. This serves as evidence, that the origin of the hysteretic charge/discharge behaviour is not a formation of a core and a shell. Thus these models are not suitable to find suggestions on how to influence this behaviour.

## 2.3 Discrete Many-Particles Model

The origin of hysteretic behaviour in the *discrete many-particles model* stems from the interaction of all particles as an ensemble, and the possibility to be charged asynchronous. This asynchrony manifests itself in the occurrence of particles with different states at the same time. We assume that the change of the loading state is slow, compared to the interaction speed of the ensemble of particles in which it evolves into thermodynamical equilibrium. Thus we assume *quasi static loading*. At the end of this section we will discuss advantages and shortcomings of this model and motivate the investigation of the *continuous many particle model* which we analyse in the following chapters. The model described in this section is also described in detail by Dreyer, Gaberscek, Guhlke, Huth, Jamnik and Moskon in [14] as well as by Dreyer, Guhlke and Huth in [12]. A model of a similar type has been analysed in the context of mechanical deformation of a bistable chain of springs by Puglisi and Truskinovsky [35] and by Mielke and Truskinovsky in [30].

First we will discuss an illustrative analogon. The principles of connected storage particles fit in the framework of the work by Dreyer, Müller and Strehlow [10] on interconnected rubber balloons. There the characteristic of pressure versus air volume in a balloon is given by a nonmonotone function. This is a result of rubber having a nonmonotone stress strain relation at room temperature. It can be experienced with a




(a) relation of the pressure in a balloon to the degree of inflation, taken from [31]

(b) prototype of a nonmonotone potential

Figure 2.5: Non-monotone material behaviour.

common balloon. The force a person has to apply while inflating a balloon is at first high and then, after passing a critical state of inflation, gets lower again. This manifests in a local maximum for a small filling degree in Figure 2.5(a), followed by a decreasing part of the curve. Later on when the balloon gets bigger, the force necessary for inflation gets higher again, which is described by the curve being increasing after a local minimum in the same figure.

The main phenomenon which we want to highlight is that the interval in which the pressure curve is decreasing, is an unstable region for a system of connected balloons. Imagine two balloons $A$ and $B$ having the same degree of inflation $x_A$ and $x_B$. The value $x_A = x_B$ shall be in the decreasing region of the pressure curve. Furthermore both balloons are connected and can exchange air, but no air is exchanged with the outside, such that the overall air volume $x_A + x_B$ stays constant. Now assume that this state is slightly perturbed such that $x_A < x_B$, but still both $x_A$ and $x_B$ are in the region where the pressure decreases if the air volume increases. Thus for the pressure $p$ there holds $p(x_A) > p(x_B)$. Since the balloons are connected, this means that the smaller balloon, A, has higher pressure than B and consequently inflates the larger balloon, because of the pressure difference. By this the pressure difference gets even bigger. As a result for many interconnected balloons, no stable state exists in which more than one would be in this region.

We consider the similar situation of an electrode which stores lithium in an $FePO_4$ lattice as described in Section 2.1. The storage and extraction of lithium takes place in a large number of $FePO_4$ particles at the same time. In contrast to the core shell model we assume that the particles are homogeneous at all times. This is motivated by experimental observations of Delmas et al. [8]. They observe that while loading iron phosphate particles with lithium these do not form a core shell structure, but instead lithium rich and lithium poor particles coexist. In the same work it is shown, that once the lithium intercalation starts in a particle, further intercalation gets easier in this

particle compared to other lithium poor particles. Additionally in a paper by Wagemaker et al. [41], experiments show that particles with low and high lithium concentration can coexist next to each other in one electrode. The authors used another storage material for the lithium acceptor electrode, namely anatase ($TiO_2$). Thus this phenomenon is not restricted to $FePO_4$ electrodes.

We will deduce observations for the model described in this section which are similar to the observation described for the balloons. The behaviour for the ensemble of $FePO_4$ particles relies only on assumptions on the potential of a single storage particle. In our model the potential is comparable with the pressure versus air volume characteristics for a balloon. The potential is related to the voltage of a $FePO_4$ particle measured against the lithium electrode. It is the derivative of an energy. The modelling of this energy from physical principles can be found in [12]. It can be expressed as a function of the degree of filling with lithium of the particle, $x \in [0,1]$. A value of $x = 1$ means that the particle is filled with the maximal possible amount of lithium, and $x = 0$ means that the particle is empty. The filling degree can also be seen as a measure of the electrical loading state of the single particle in terms of electrical energy. Here $x = 0$ means fully charged and $x = 1$ fully discharged. In the sequel we will therefore not distinguish between loading state and degree of lithium filling since either one can be expressed by the other quantity.

For the energy $\psi : [0,1] \to \mathbb{R}$ and the chemical potential $\psi'$ describing a single particle, we assume that it is qualitatively as depicted in Figure 2.5(b). To be precise we assume the following.

**Assumption 2.3.1.** *We assume that $\psi$ and $\psi'$ satisfy (A1) to (A3).*

*(A1)* $\psi \in C^2(]0,1[;\mathbb{R}) \cap C([0,1];\mathbb{R})$.

*(A2) There exist $x_*, x^* \in ]0,1[$ with $x_* < x^*$, so that $\psi'$ is strictly decreasing in the interval $[x_*, x^*]$ and strictly increasing elsewhere. Hence $\psi''(x) = 0 \Leftrightarrow x \in \{x_*; x^*\}$.*

*(A3) The potential $\psi'$ satisfies $\lim_{s\to 0} \psi'(s) < \psi'(x)$ for all $x \in ]0,1[$ and $\psi'(x) < \lim_{s\to 1} \psi'(s)$ for all $x \in ]0,1[$.*

We also need to define the values $x_{**}$ and $x^{**}$. They denote the maximal boundary of the set $S = ]0, x_{**}[ \ \cup \ ]x^{**}, 1[$, such that each value in $\{y \in \mathbb{R} : \ \exists x \in S \ , y = \psi'(x)\}$ has only one preimage on the whole unit interval under the mapping $\psi'$.

The loading state of the battery, $\ell$, is the mean value of the relative loading states of all particles

$$\ell = \frac{1}{N} \sum_{k=1}^{N} x_k. \tag{2.2}$$

We prescribe the global loading state $\ell \in [0,1]$ for the whole battery at all times, and hence (2.2) gives an extra condition on the $N$-tuple $x$.

As a result of the modelling process in [12], the so called available free energy $A$ of the system of $N$ connected storage particles can be expressed as a function in $x =$

$(x_1, \cdots, x_N) \in \mathbb{R}^N$, where $x_k \in [0,1]$ is the relative loading state of the $k$-th particle. It is the sum of the energies of the single particles,

$$A(x) := \frac{1}{N} \sum_{k=1}^{N} \psi(x_k) \tag{2.3}$$

and this $\psi$ satisfies the assumptions (A1)-(A3). Equilibria of the system at hand are identified as minima of the energy $A$. We also define the mean potential. The voltage of the battery is linearly depending on the mean potential

$$\langle \psi' \rangle(x) := \frac{1}{N} \sum_{k=1}^{N} \psi'(x_k). \tag{2.4}$$

We assume, that the process of charging is relatively slow compared to the speed of the system to move to equilibria states. This means that the rate of change of $\ell$ is much smaller than the speed in which the system moves to (local) minima of the free available energy $A$. Therefore we assume, that at any time the system is in an equilibrium state, which means a state which is a local minimum of $A$. This is called quasi static loading.

At any time $t$ we have a prescribed loading state $\ell(t)$ and so we are interested in minimisers of $A$ among all possible $N$-tuples, representing this loading state according to (2.2). For this we define the manifold $M_r$ as

$$M_r := \Big\{ x \in [0,1]^N : r = \frac{1}{N} \sum_{k=1}^{N} x_k \Big\}.$$

We can now deduce properties for minimisers of $A$ on manifolds $M_\ell \subset \mathbb{R}^N$ by exploiting the properties of the energy given in Assumption 2.3.1. The first one is that if the loading state $\ell \in ]0,1[$ is away from the boundary of the unit interval, then the relative loading states of $x_k$ must stay away from the boundary as well.

**Lemma 2.3.2.** *Assume that $\ell \in ]0,1[$ and that $x$ is local minimum of $A$ on $M_\ell$. The Assumption 2.3.1-(A3) implies that no local minimum of $A$ on the manifold $M_\ell$ lies on the boundary of $[0,1]^N$.*

*Proof.* We prove this by contradiction. Say that there is a local minimiser $x \in [0,1]^N$, such that for an index $k_1$, $x_{k_1} = 0$. Taking in account the definition of $\ell$ in (2.2) we deduce by $\ell > 0$ that there must exist another index $k_2$ such that $x_{k_2} > 0$. Define the continuous function $h : [0, \varepsilon] \to M_\ell$ as

$$\Big(h(s)\Big)_j := \begin{cases} s \ , & \text{if } j = k_1 \ . \\ x_{k_2} - s \ , & \text{if } j = k_2 \ . \\ x_j \ , & \text{otherwise} \ . \end{cases}$$

Obviously $h(0) = x$ and for $s$ small enough $h(s) \in M_r$. We observe the continuous change in the energy along the trajectory $h$ as $s$ approaches zero. There is a $c \in \mathbb{R}$ such

that

$$A(h(s)) = c + \psi(s) + \psi(x_{k_2} - s) \ , \qquad \frac{\mathrm{d}}{\mathrm{d}s} A(h(s)) = \psi'(s) - \psi'(x_{k_2} - s) \ .$$

Exploiting (A3) in Assumption 2.3.1 leads to

$$\lim_{s \to 0} \frac{\mathrm{d}A(h(s))}{\mathrm{d}s} = \lim_{s \to 0} \psi'(s) - \psi'(x_{k_2}) < 0 \ ,$$

which means that $A(h(s))$ strictly increases while $s \searrow 0$ and $h(s) \to x$. This contradicts with $x$ being a local minima.

In the second case which we should consider, $x$ is a local minimum of $A$ on $M_\ell$ and there is one index $k_1$, such that $x_{k_1} = 1$. The contradiction can then be constructed analogously, and so we have proven this lemma. □

As a consequence of this lemma we can employ simple optimality criteria to further investigate the minima. The results are formulated in the next lemma.

**Lemma 2.3.3.** *Let Assumption 2.3.1 hold. We define $A$ as in (2.3). Fix $N \in \mathbb{N}$ and $\ell \in ]0,1[$. If $x \in \mathbb{R}^N$ is a minimiser of $A$ on $M_\ell$, then there exists a positive constant $\delta_N > 0$ such that*

$$\psi'(x_k) = \langle \psi' \rangle(x) \ , \qquad \text{for all} \quad k = 1, 2, \cdots, N \ , \tag{2.5}$$

$$x_k \in ]0, x_* + \delta_N[ \ \cup \ ]x^* - \delta_N, 1[ \ , \qquad \text{for all} \quad k = 1, 2, \cdots, N \ , \tag{2.6}$$

*and all particles with mole fraction $x_k \in [0, x_*]$ have the same value as well as all particles with mole fraction $x_k \in [x^*, 1]$ have the same value.*

*For the sequence $\{\delta_N\}_{N \in \mathbb{N}}$ belonging to all possible choices of dimensions $N$ we know that $\delta_N \to 0$ as $N \to \infty$.*

*Furthermore at most one particle state is in the intermediate region $]x_*, x^*[$. By this we mean that there exists at most one $k_* \in \{1; 2; \cdots; N\}$ such that $x_{k_*} \in ]x_*, x^*[$.*

*Proof.* In order to examine minima on $M_\ell \subset \mathbb{R}^N$ we parametrise this manifold by $\Psi$,

$$M_\ell^0 := \Big\{ y \in \mathbb{R}^{N-1} : y_k \in [0,1] \text{ and } (N\ell - \sum_k y_k) \in [0,1] \Big\} \ ,$$

$$\Psi : M_\ell^0 \to M_\ell \ , \qquad \Psi(y) \mapsto x = (y_1, \cdots, y_{N-1}, N\ell - \sum_k y_k) \ .$$

This allows to equivalently consider local minima of

$$A_0 : M_\ell^0 \to \mathbb{R} \ , \qquad A_0(y) := A(\Psi(y)) = \frac{1}{N} \psi(N\ell - \sum_k y_k) + \frac{1}{N} \sum_{k=1}^{N-1} \psi(y_k) \ .$$

By Lemma 2.3.2 we know that local minima of $A$ are not on the boundary of $M_\ell$ and so any local minima $y_*$ of $A_0$ on $M_\ell^0$ must also be inside of $M_\ell^0$. It is thus a necessary

condition that the derivative of $A_0$ vanishes at $y_*$ and its second derivative is positive semidefinite. Setting the derivative of $A_0$ to zero immediately gives (2.5).

We can treat the simple case $\ell < x_{**}$ or $\ell > x^{**}$ in advance, because then $x_k = \ell$ for all $k$ and thus (2.6) holds. Indeed assume $x$ is a minimiser of $A$ on $M_\ell$ with $\ell = \frac{1}{N}\sum_k x_k < x_{**}$. Clearly for at least one index $j$ we must have that $x_j < x_{**}$. Hence by the definition of $x_{**}$ on page 13 we deduce by (2.5) that $x_k = \ell$ for all $k$. For $\ell > x^{**}$ the same reasoning holds mutatis mutandis.

In the second case where $\ell \in [x_{**}, x^{**}]$, we conclude for minimisers $x$ of $A$ on $M_\ell$ by (2.5) that $x_k \in [x_{**}, x^{**}]$ for all $k$. Hence it follows from the continuity of $\psi'$ and $\psi''$ that there exists a constant $c_0$ such that for any $N$ and any minimiser $x \in M_\ell$

$$|\psi''(x_k)| \le c_0 \qquad \text{for all } k\ . \tag{2.7}$$

In order to prove (2.6) we consider an index $k_* \in \{1, \dots, N)$ such that $x_{k_*} \in ]x_*, x^*[$, which implies $\psi''(x_{k_*}) \le 0$. Next, we define the vector $v \in \mathbb{R}^{N-1}$ as

(a) if $k_* < N$, $v := e_{k_*} - \frac{1}{N-1}\sum_{i=1, i\neq k_*}^{N-1} e_i$,

(b) if $k_* = N$, $v := \frac{1}{N-1}\sum_{i=1}^{N-1} e_i$,

where we denote by $e_i \in \mathbb{R}^{N-1}$ the $i$-th unit vector, which has an entry 1 at the $i$-th position and zero entries otherwise. We look at the Hessian $H$ of $A_0$,

$$H(y) = D^2 A_0(y) = \frac{1}{N}\Big( \begin{pmatrix} \psi''(x_1) & & 0 \\ & \ddots & \\ 0 & & \psi''(x_{N-1}) \end{pmatrix} + \psi''(x_N) \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix} \Big)\ .$$

It must be positive semi-definite at the critical points $y = (x_1, \cdots, x_{N-1})$. In both cases (a) and (b), we conclude

$$0 \le v^T H(y) v = \psi''(x_{k_*}) + \frac{1}{(N-1)^2} \sum_{i=1, i\neq k_*}^{N-1} \psi''(x^i)\ . \tag{2.8}$$

Thus, due to (2.7), we have

$$0 \le |\psi''(x_{k_*})| \le \frac{\frac{1}{N-1}\sum_{i=1, i\neq k_*}^{N-1} \psi''(x^i)}{N-1} \le \frac{c_0}{N-1}\ . \tag{2.9}$$

This means for sufficiently large $N$, that $\psi''(x_{k_*})$ is close to zero. By Assumption 2.3.1 $\psi''$ is continuous and vanishes only at positions $x_*$ and $x^*$. Hence, we deduce from (2.9) that $x_{k_*} \in ]x_* - \delta_N, x_* + \delta_N[\ \cup\ ]x^* - \delta_N, x^* + \delta_N[$, and furthermore that $\delta_N \to 0$ as $N \to \infty$.

What remains is to show that only one mole fraction is in the intermediate region $]x_*, x^*[$. Again we only consider the nontrivial case $\ell \in [x_{**}, x^{**}]$. Say that two indices $k_1, k_2$ exist such that $x_{k_1}, x_{k_2} \in ]x_*, x^*[$. This implies $\psi''(x_{k_1}), \psi''(x_{k_2}) < 0$ and we define a vector $v$ as

(a) if $k_1, k_2 < N$ define $v := e_{k_2} - e_{k_1}$ then we have $v^T H(x) v < 0$.

(b) if $k_1 < N, k_2 = N$ define $v := e_{k_{k_1}}$ then we have $v^T H(x) v < 0$.

Then we can not have a local minimum, since $H$ is not positive semidefinite. □

The number of nano-sized storage particles in a battery electrode is in ranges of more then $10^{10}$ particles, see [14]. Taking Lemma 2.3.2 into account, this makes following assumption reasonable. Therein we assume that no particle in the whole ensemble is of intermediate loading state.

**Assumption 2.3.4.** *We consider only minimiser $x = (x_1, \cdots, x_N)$ of $A$ on $M_\ell$ which satisfy (A4).*

*(A4) There shall be no index $k \in \{1, 2, ..., N\}$ such that a mole fraction $x_k$ is in the intermediate region $]x_*, x^*[$.*

With this assumption, every equilibrium state $x \in M_\ell$ consists of two kinds of particles. We interpret this as the formation of two phases. Each phase consists of particles having the same filling degree, which is either low in the $\alpha$-phase or high in the $\beta$-phase. The two filling degrees are $x_\alpha \in [0, x_*]$ and $x_\beta \in [x^*, 1]$. Due to the non-monotone nature of $\psi'$ and the non-convexity of $A$, there are in general many local minima of $A$ on the manifold $M_\ell$. The local minima of $A$ satisfying (A4), which we call shortly stable states, differ by the number of particles in each phase and by the phase properties $x_\alpha$ and $x_\beta$.

In order to be independent of the exact number and ordering of particles, we introduce the phase fraction of a stable state $x \in [0, 1]^N$ as

$$\lambda(x) := \frac{\#\{k : x_k \in [x^*, 1]\}}{N} . \tag{2.10}$$

We consider $\lambda$ as a macroscopic quantity of the system. If Assumption 2.3.4 holds and if we consider only minima of $A$, then knowing the two macroscopic quantities $\lambda(x)$ and $\ell(x)$ uniquely determines the two loading states of the $\alpha$ and $\beta$ phase. This is formulated in the following lemma, where we exploit Assumption 2.3.4 and Lemma 2.3.3.

**Lemma 2.3.5.** *Let the Assumptions 2.3.1 and 2.3.4 be satisfied. Depending on $\lambda \in [0, 1]$ we define the interval $I_\lambda := [\ell_\lambda^-, \ell_\lambda^+]$ with $\ell_\lambda^- := \lambda x^* + (1-\lambda) x_{**}$ and $\ell_\lambda^+ := \lambda x^{**} + (1-\lambda) x_*$. There exist two unique functions*

$$x_\alpha : \{(\lambda, \ell) \in \mathbb{R}^2 : \lambda \in [0, 1], \ell \in I_\lambda\} \to [x_{**}, x_*] ,$$
$$x_\beta : \{(\lambda, \ell) \in \mathbb{R}^2 : \lambda \in [0, 1], \ell \in I_\lambda\} \to [x^*, x^{**}] ,$$

*which satisfy for all $\ell$ and $\lambda$,*

$$\ell = \lambda \; x_\beta(\lambda, \ell) \; + \; (1 - \lambda) \; x_\alpha(\lambda, \ell) , \tag{2.11}$$
$$\psi'(x_\alpha(\lambda, \ell)) = \psi'(x_\beta(\lambda, \ell)) . \tag{2.12}$$

*Both functions $x_\alpha$ and $x_\beta$ are continuous in $\ell$ and $\lambda$. The potential $\psi'(x_\alpha(\lambda, \ell))$ is even equicontinuous in $\lambda$.*

*Proof.* The case when $\lambda = 1$ or $\lambda = 0$ is trivial, and we therefore treat only the case $\lambda \in ]0,1[$. We will exploit that this allows us to consider values of $\ell, x_\alpha, x_\beta$ on the compact interval $[x_{**}, x^{**}]$ where the function $\psi$ and its derivatives are bounded. In order to show the existence of functions satisfying (2.11) and (2.12), we define

$$\begin{aligned} g_0 :&[\mu_*, \mu^*] \to [x_{**}, x_*], & g_0 :&= \big(\psi'|_{[x_{**}, x_*]}\big)^{-1}, \\ g_1 :&[\mu_*, \mu^*] \to [x^*, x^{**}], & g_1 :&= \big(\psi'|_{[x^*, x^{**}]}\big)^{-1}, \end{aligned}$$

as the inverses of the two monotone increasing branches of $\psi'$. Furthermore we define

$$g_\lambda :[\mu_*, \mu^*] \to [x_{**}, x^{**}] \ , \qquad g_\lambda :=\lambda g_1 + (1-\lambda) g_0. \tag{2.13}$$

Recall that the functions $g_{0,1}$ are strictly monotone increasing. This implies that $g_\lambda$ is strictly monotone increasing. Therefore $g_\lambda([\mu_*, \mu^*]) = I_\lambda$ and the function

$$\mu_\lambda(z) := g_\lambda^{-1}(z) \tag{2.14}$$

is well defined and continuous on $I_\lambda$. Now we define $x_\alpha$ and $x_\beta$ as

$$x_\alpha(\lambda, \ell) := g_0(\mu_\lambda(\ell)) \ , \qquad x_\beta(\lambda, \ell) := g_1(\mu_\lambda(\ell)) \ ,$$

which are continuous and well defined on $I_\lambda$. Note that (2.11) and (2.12) are valid by construction, and so this is a solution. The uniqueness follows from (2.11) and (2.12). If we solve equation (2.11) for $x_\alpha$ and insert $x_\alpha = (\ell - \lambda x_\beta)/(1-\lambda)$ into (2.12), we may define

$$G(x_\beta) := \psi'(x_\beta) - \psi'\Big(\frac{\ell - \lambda x_\beta}{1-\lambda}\Big) \ . \tag{2.15}$$

Any Solution of (2.11)-(2.12) must also give a root of $G$. The function $\psi'$ is strictly monotone increasing on the allowed sets $x_\alpha \in [x_{**}, x_*]$ and $x_\beta \in [x^*, x^{**}]$, hence $G$ is strictly monotone increasing in $x_\beta$ as well. Thus $G$ has a unique root on the allowed set and the above constructed functions $x_\alpha$ and $x_\beta$ are the only possible solutions.

In order to show the desired continuity of the functions in $\lambda$ we look at the continuity of $\mu_\lambda(\ell)$ when varying $\lambda$ for fixed $\ell$. By the assumed continuity of $\psi$ we know that there is a constant $k$ such that

$$\max_{x \in [x_{**}, x_*] \ \cup \ [x^*, x^{**}]} |\psi''(x)| \leq k \ .$$

Therefore $g_0'(x) \geq 1/k$ on $[x_{**}, x_*]$ and $g_1'(x) \geq 1/k$ on $[x^*, x^{**}]$. As constructed this results in $g_\lambda' \geq 1/k$ and hence

$$\max_{\lambda \in [0,1], x \in I_\lambda} |\mu_\lambda'| \leq k \ . \tag{2.16}$$

Before we use this estimate we need a further ingredient. For any small $\delta \in \mathbb{R}$ we deduce

from (2.13) for $s = \mu_\lambda(\ell)$,

$$g_{\lambda+\delta}(s) = \underbrace{\lambda g_1(s) + (1-\lambda) g_0(s)}_{=\ell} + \underbrace{\delta\big(g_1(s) - g_0(s)\big)}_{:=\hat{\delta}} \quad \Rightarrow \quad \mu_\lambda(\ell) = \mu_{\lambda+\delta}(\ell + \hat{\delta}) \ . \tag{2.17}$$

We know the range of $g_0$ and $g_1$ and so we conclude $\hat{\delta} \leq \delta |x^{**} - x_{**}|$. Using this estimate for $\hat{\delta}$ together with (2.16) and (2.17) we get for the difference

$$\mu_\lambda(\ell) - \mu_{\lambda+\delta}(\ell) = \underbrace{\mu_\lambda(\ell) - \mu_{\lambda+\delta}(\ell+\hat{\delta})}_{=0, \text{ see (2.17)}} + \int_\ell^{\ell+\hat{\delta}} \mu'_{\lambda+\delta}(s)\,\mathrm{d}s| \tag{2.18}$$

$$\leq k\hat{\delta} \leq k\delta |x^{**} - x_{**}|. \tag{2.19}$$

This gives Lipschitz continuity in $\lambda$ and hence equicontinuity of $\psi'(x_\alpha(\lambda, \ell)) = \mu_\lambda(\ell)$ as stated in the lemma. □

In summary if we assume that $x$ is a stable state and if Assumption 2.3.1 and 2.3.4 hold, then knowing two of the following three quantities is enough to uniquely determine the state $x \in \mathbb{R}^N$ up to the order of particles:

- the macroscopic potential of the System $\langle \psi' \rangle(x) = \psi'(x_\alpha) = \psi'(x_\beta)$, which is equivalent to knowing the loading state of all $\alpha$- and $\beta$-phase particles $x_\alpha$ and $x_\beta$,
- the loading state of the whole system $\ell$, see (2.2),
- the phase fraction $\lambda$, see (2.10).

Note that all of them can be seen as macroscopic quantities. Each can be expressed as a continuous function of the two others. Knowing this allows us to examine possible equilibria of the whole range of macroscopic loading states in one figure, namely Figure 2.6.

In Figure 2.6 we see all stable states of a system of $N = 10$ particles equipped with the same characteristic potential $\psi'$ for each particle as depicted in Figure 2.5(b). We want to comment on the qualitative behaviour for the stable states and therefore we do not give exact units on the graph, nor do we give the exact formula for the potential. Note that the stable states form disjunct lines. Each of them represents a set of stable states having the same phase fraction $\lambda(x)$. We call this set a branch or in short $B_{\lambda(x)}$. Since we consider a system of 10 particles, only a finite number of phase fractions are possible, namely $\lambda = k/10$ with $k = 0, \cdots, 10$, which explains the occurrence of different branches.

Under the assumption of quasi static loading we postulate that the system must be in a stable state for all times. We also postulate that the system stays on a branch belonging to a fixed phase fraction $\lambda$ when possible. This means that $\lambda(t) = \lambda_0$ until $t$ reaches a critical time. As can be seen in Figure 2.6, one branch does not cover all possible values for loading states $\ell$, but only a small interval $[\ell^-(\lambda), \ell^+(\lambda)]$. So when the loading state $\ell(t)$ increases over $\ell^+(\lambda)$ in time, the system must leave the branch $B_\lambda$ and

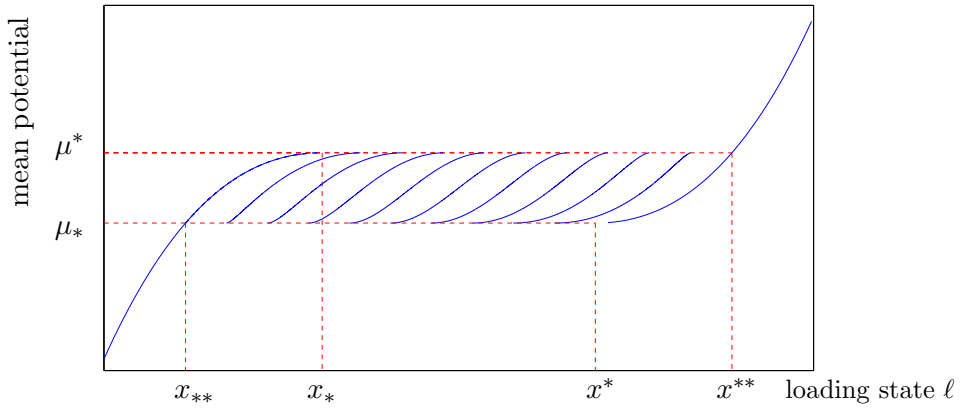



Figure 2.6: Possible stable states marked with solid lines in a plot, showing the mean potential $\langle \psi' \rangle$ on the vertical axis versus the global loading state $\ell$ on the horizontal axis.

jump onto another branch. We constitute that the system jumps on the next closest branch belonging to the phase fraction $\lambda_1 = \lambda_0 + 1/n$. While unloading, we assert the same behaviour on the lower end, $\ell^-(\lambda)$, of each branch.

Let the global loading state $\ell$ be a given continuous and differentiable function of time. The dynamic behaviour which we constitute, can be condensated in the following axioms:

(AX1) The systems evolution starts within a stable state, i.e. it can be described by a pair $(\ell(t_0), \lambda(t_0))$ and belongs to a branch $B_\lambda$ such that $\ell(t_0) \in I_{\lambda(t_0)}$.

(AX2) While the loading state $\ell$ changes in time, the system stays, if possible, on the branch $B_\lambda$ where it was.

(AX3) When the system state belongs to the branch $B_\lambda$ and the loading state $\ell$ increases in time over the maximal possible loading state $\ell^+(\lambda)$ belonging to this phase fraction $\lambda$, or decreases under the minimal possible loading state $\ell^-(\lambda)$, then the phase fraction $\lambda$ is raised or lowered by $1/N$. This means that the system state jumps from one branch to the other one, or that one particle of the ensemble changes its phase.

This behaviour is depicted in Figure 2.7. As proven by the equicontinuity of the potential in Lemma 2.3.5, these axioms result in the limit for infinite particle number, $N \to \infty$, in a smooth behaviour, which can be seen in Figure 2.8. Thus we are able to reproduce the hysteretic behaviour as experienced in measurements while loading and unloading.

The limiting behaviour can be described with an ODE for the macroscopic quantity $\lambda$ depending on the given function of time $\ell$. We define $I_\lambda, \ell^-_\lambda$ and $\ell^+_\lambda$ as in the proof of Lemma 2.3.5. Say that on a time interval $S = [t_0, T]$ the loading state is given as a continuous and differentiable function $\ell : S \to [0, 1]$ as well as admissible $\lambda(t_0)$ such that

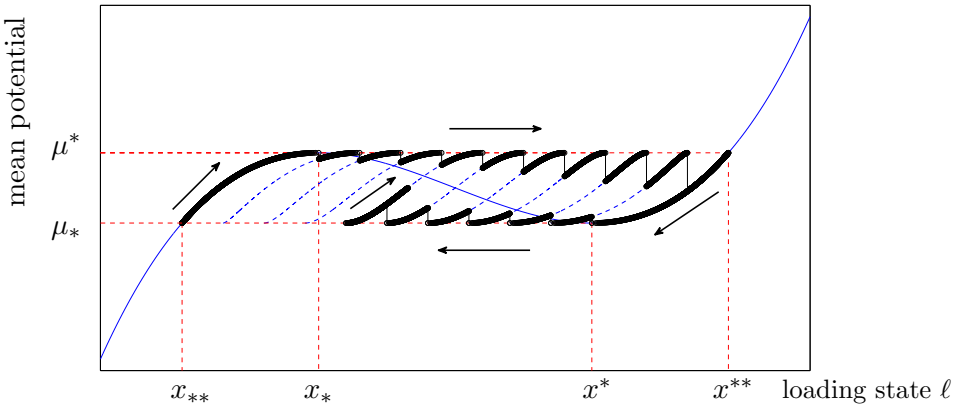



Figure 2.7: Quasistatic behaviour of a 10 particle system while loading and unloading, we see the evolution of the mean potential versus the loading state (thick solid ), the shape of the original single particle potential (thin solid), stable paths (dashed).

$\ell(t_0) \in I_{\lambda(t_0)}$. Using Lemma 2.3.5 this allows to determine the mole fractions in the two phases and the potential of the system. The limiting behaviour then is

$$\frac{\mathrm{d}\lambda}{\mathrm{d}t} = \begin{cases} 0 \ , & \text{if } \ell(t) \in I_\lambda \ . \\ \frac{1}{x^{**}-x_*} \ \frac{\mathrm{d}}{\mathrm{d}t}\ell \ , & \text{if } \ell(t) = \ell_\lambda^+ \ . \\ \frac{1}{x^{*}-x_{**}} \ \frac{\mathrm{d}}{\mathrm{d}t}\ell \ , & \text{if } \ell(t) = \ell_\lambda^- \ . \end{cases} \tag{2.20}$$

It can also describe so called inner loops as seen in Figure 2.8, which can be experienced in experiments, see Srinivasan and Newman [39]. The core shell model was not able to reflect these inner loops.

So why do we need another model? The constitutive axiom (AX3), which states that only one particle changes its phase if necessary, seems to be the backbone of this model. An explanation of this axiom as a consequence of some physical necessity is thus desirable. Especially if ones aim is to understand and even change the (dis)charging characteristic of a battery. Remember that experiments of inflating connected balloons are used as analogon to the model at hand. There for small speed of inflation the balloons behave as described with the axiom (AX3). One by one the balloons change their phase from small to big balloons. On the other hand, for a fast speed of inflation, it happens that more than one balloon changes its phase from small to big at a time. This also motivates, that this axiom should be questioned more, especially when one wants to investigate and improve the behaviour while (dis)charging a battery fast.

Similar doubts are raised when considering a naive expansion of the above model to non-quasi-static loading. Say that $y : [0,T] \to \mathbb{R}^N$ is the euclidian gradient flow to the energy $A$ in (2.3) combined with an extra term that guarantees the constraint $\frac{1}{N}\sum_k y_k(t) = \ell(t)$. This results for given constant $\tau > 0$ and function $\ell$ into the evolution

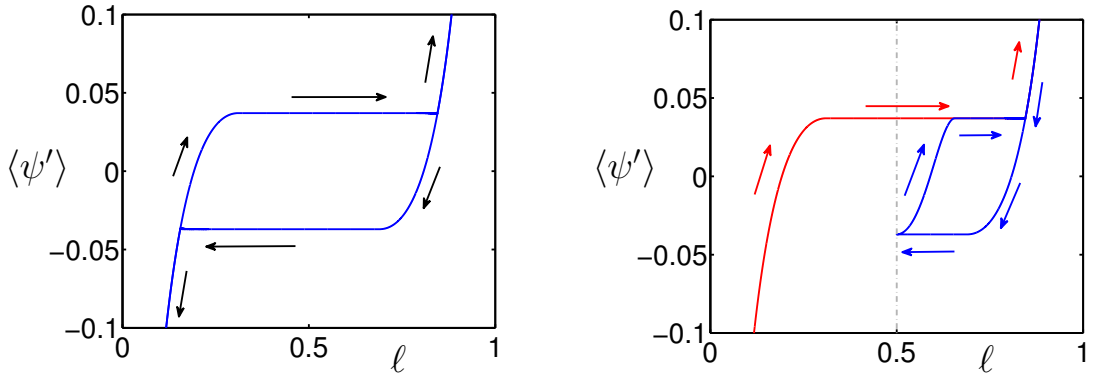



Figure 2.8: Dynamic behaviour of discrete model for particle number $N \to \infty$, when observing the evolution of the mean potential $\langle \psi' \rangle(t) := \psi'(x_\alpha(t)) = \psi'(x_\beta(t))$ against the time dependent loading state $\ell$, taken from [12] .

equation

$$\dot{y}_k(t) = -\tau \psi'(y_k(t)) + \tau \frac{\sum_{i=1}^{N} \psi'(y_i(t))}{N} + \dot{\ell}(t), \qquad \text{for all } k = 1, \cdots, N \ . \tag{2.21}$$

In a short time the components of the solution to (2.21), $y_k$, have almost identical values. They show no behaviour similar to the assumed jumping in Axiom (AX3). A simulation of this behaviour for $\tau = 10^4$ is depicted in Figure 2.9. The effect that all components behave (approximately) the same even when crossing the unstable region is called delayed bifurcation effect and was also seen by Mielke and Truskinovsky in [35]. In their work Mielke and Truskinovsky introduce a bias in the characteristics or material behaviour of the single particles (in their work springs), in order to overcome this phenomenon.

We modify the ODE (2.21) in the spirit of the following model in this work by adding formally a noise $X$.

$$\dot{y}_k(t) = -\tau \psi'(y_k(t)) + \tau \frac{\sum_{i=1}^{N} \psi'(y_i(t))}{N} + \dot{\ell}(t) \ + \ X_k(t), \quad \text{for all } k = 1, \cdots, N \ . \tag{2.22}$$

The noise $X$ shall not influence the mean value of $y$, which means that $X(t) \in \mathbb{R}^N$ shall be mean free at all times. It is certainly not in the scope of this work to enter the field of stochastic differential equations. Nevertheless we want to give an impression of how the influence of a noise $X$ in (2.22) might influence the solutions. For this we show a naive numerical simulation of (2.22) in Figure 2.10. This can be seen as the solution along a possible path. At every numerical time step $j$ we created a vector $Z \in \mathbb{R}^N$ of $N$ (pseudo) random numbers from a uniform distribution on the open interval $]0, 1[$. Then we compute the value for $X = (X_1, X_2, \cdots, X_N)$ as

$$X_k = \eta\Big(Z_k - \frac{\sum_{i=1}^{N} Z_i}{N}\Big) \ ,$$

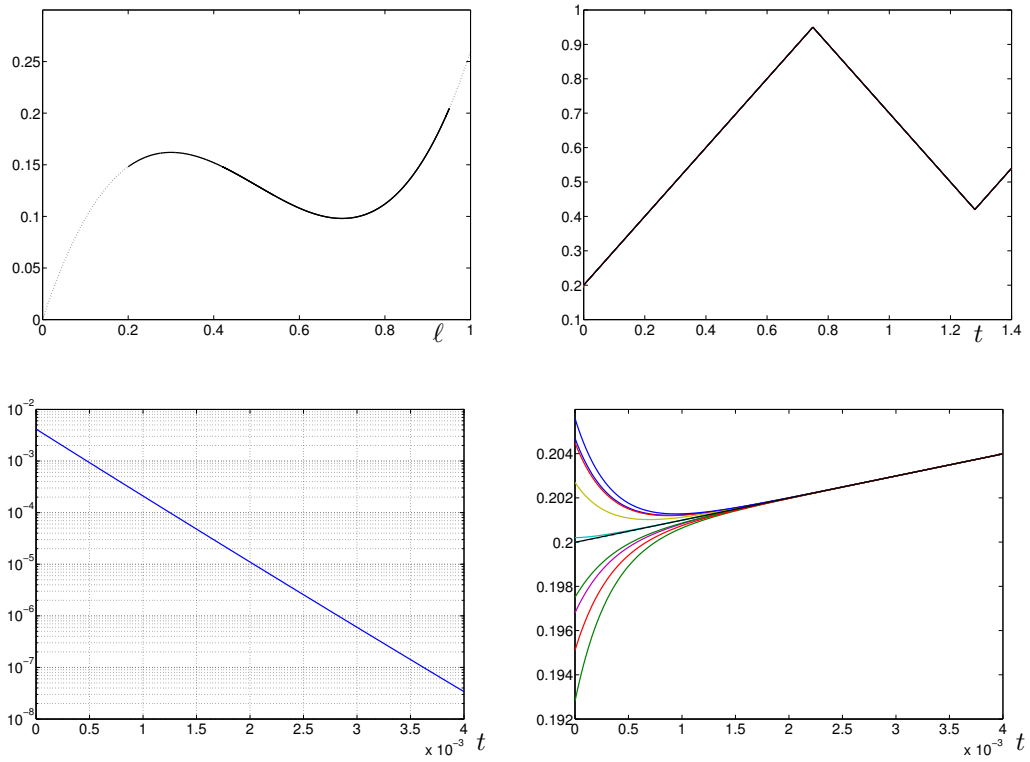



Figure 2.9: Simulation of ODE (2.21) with $\tau = 10^4$, which is a gradient flow of a many particle system with no noise (10 particles), upper left: evolution of $\langle\psi'\rangle(t) := \frac{1}{N}\sum_k \psi'(k)$ vs $\ell(t)$ (solid), and the function $\psi'$ (dotted), lower left: evolution of the variance of the components $y_k$, upper and lower right: evolution of the components (particles) $y_k$ (solid) and the given function of time $\ell$ (dashed).

where the constant $\eta > 0$ regulates the strength of the noise in comparison to the strength of the drift of the gradient flow $\tau$. Our simulation with constants $\tau = 10^4$ and $\eta = 10$ is shown in Figure 2.10. Note the astonishing agreement with the evolution in Figure 2.7. The experience gathered by numerical simulations with various values for $\tau$ and $\nu$ shows, that the relation of $\tau$ and $\nu$ has a great influence on the qualitative behaviour of the ensemble. We can see the occurrence of the delayed bifurcation effect (all particles behave the same as in Figure 2.9) on the one hand or on the other hand the jumping of one particle after the other as in Figure 2.10. Even behaviour in between is possible, where the ensemble crosses the unstable region not one after the other but in groups. We end our visit to the stochastic differential equations here, with a concluding remark. Motivated from Figure 2.10 we believe that equation (2.22) can be transformed into the nonlinear Fokker–Planck equation (3.9) which is the subject of the following chapters.

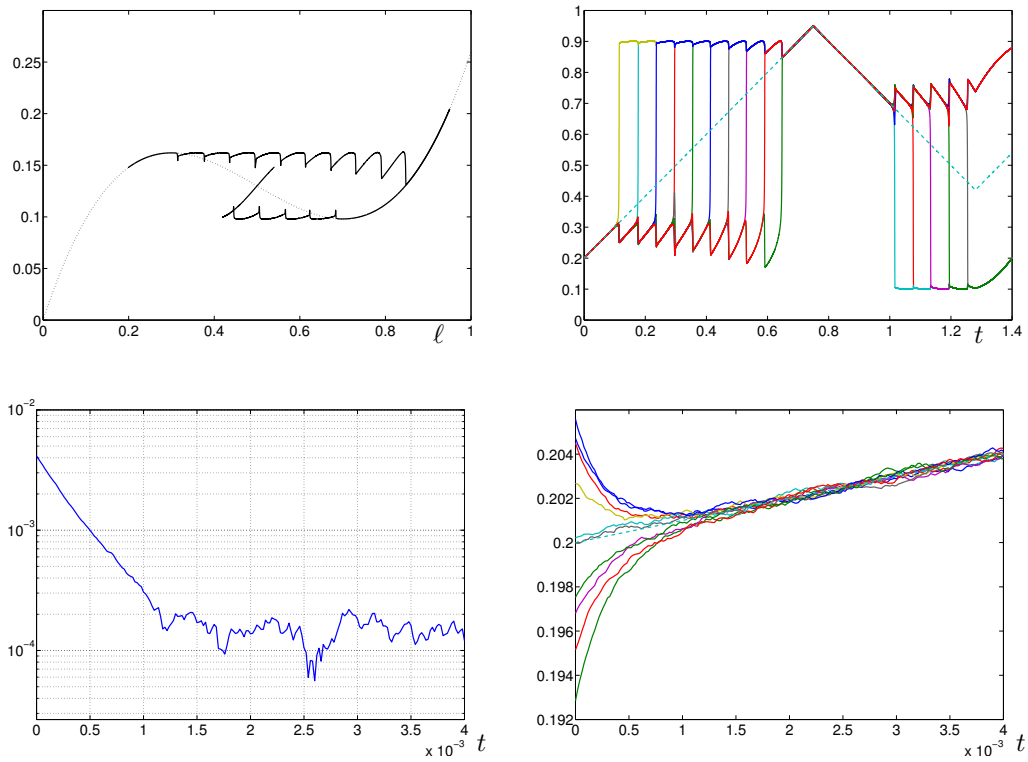



Figure 2.10: Simulated gradient flow of a many particle system with noise (10 particles), see evolution equation (2.22), upper left: evolution of $\langle \psi' \rangle(t) := \frac{1}{N} \sum_k \psi'(k)$ vs $\ell(t)$ (solid), and the function $\psi'$ (dotted), lower left: evolution of the variance of the components $y_k$, upper and lower right: evolution of the components (particles) $y_k$ (solid) and the given function of time $\ell$ (dashed).

# 3 Deriving a Fokker–Planck Equation to Model a Lithium Ion Battery

The model presented here is the successor of the *discrete many particles model* in Section 2.3. It is also the model of our main interest and will be analysed and discussed in the following chapters. Like the earlier models described in Chapter 2 this model describes only the intercalation of lithium atoms in an electrode consisting of many nanosized $FePO_4$-particles as described in 2.1. This chapter gives a sketch of the modelling process. For more details we refer to the detailed modelling process done by Dreyer, Guhlke and Herrmann in [11].

As in Section 2.3 we also consider the electrode as a system of connected particles. Each of them can be described by their lithium filling degree $x$. This degree can vary for each particle between $x = 0$ for empty particles, meaning no lithium atoms are intercalated into this particle, up to $x = 1$, meaning that this particle is maximally filled with lithium atoms. As in the previous chapter we will not distinguish between filling degree or electrical loading state, since each can be converted into the other. If the electrode is filled maximally with lithium atoms, then it is fully discharged, while a fully charged battery would coincide with the $FePO_4$ electrode being empty of lithium.

In contrast to the so called *discrete many particles model* in section 2.3, we do not describe the state of the system with an $N$-tuple, $(x_1, x_2, \cdots, x_N)$, having the information of all loading states for every particle. In the *continuous many particles model* we describe the system state at every time $t$ with a non-negative function $u(t) : [0, 1] \to \mathbb{R}^+$ which has mean value one. By doing so we pay tribute to the large number of nanosized particles which has values larger than $10^{10}$ in real existing batteries. This function belongs to a statistical ansatz, in which $u$ is the probability density for the distribution of the loading state $x \in [0, 1]$ of all particles. Thus the probability $P_t(\hat{x})$ of particles having a filling degree less than or equal to $\hat{x}$ at a certain time $t$, is given by

$$P_t(\hat{x}) = \int_0^{\hat{x}} u(t, x) \, \mathrm{d}x \, .$$

## 3.1 The Free Energy of the System

In order to describe the behaviour of the the underlying physical system, Dreyer, Guhlke and Herrman derive in [11] a so called total free energy $\mathcal{A}$. Its derivation depends on the assumption that the system is isothermal as well as that the kinetic energy of the system can be ignored. The energy $\mathcal{A}$ can be computed from the system state. Let $\Omega := ]0, 1[$.

The system state at time $t$ is described by a non-negative function

$$u(t) : [0,1] \to \mathbb{R}^+ \qquad \text{with} \qquad \int_\Omega u(t,x)\,\mathrm{d}x = 1\,, \tag{3.1}$$

Here and throughout this work we use $u(t)$ as shorthand for the function $\{x \to u(t,x)\}$. The global degree of filling at a time $t$ can be computed from the systems state $u(t)$ as the functional $\mathcal{C}(u(t))$, where we define $\mathcal{C} : \mathrm{L}^1(\Omega) \to \mathbb{R}$ as

$$\mathcal{C}(v) := \int_\Omega x\ v(x)\ \mathrm{d}x\,, \qquad \forall v \in \mathrm{L}^1(\Omega)\,. \tag{3.2}$$

Note that for probability distributions $v$ this can be interpreted as the expected value of $x$.

In order to write down the formula for the energy $\mathcal{A}$ we introduce the material parameters $\nu > 0$ and $\psi \in \mathrm{H}^1(\Omega)$, where $\psi$ describes the energy of a single particle according to its loading state as introduced in Section 2.3. The free energy $\mathcal{A}$ then reads

$$\mathcal{A}(t) := \int_\Omega \psi(x)u(t,x) + \nu^2 u(t,x)\log u(t,x)\,\mathrm{d}x. \tag{3.3}$$

It is composed of two contributions, whose influence is balanced by the factor $\nu$

$$\begin{aligned} &\int_\Omega \psi(x)u(t,x)\,\mathrm{d}x && \text{- sum of free energy of all particles,}\\ &\int_\Omega u(t,x)\log u(t,x)\,\mathrm{d}x && \text{- entropic mixing energy.} \end{aligned}$$

The first contribution can be seen as the continuous version of the sum of the energies of all particles in the ensemble. The second contribution gets higher, if the distribution of particles gets concentrated at a point and is lowest if $u$ represents a uniform distribution. This is a consequence of $\xi \mapsto \xi\log\xi$ being convex.

According to thermodynamic laws, see [11], the free energy must then satisfy

$$\frac{\mathrm{d}\mathcal{A}(t)}{\mathrm{d}t} - \Lambda(t)\frac{\mathrm{d}\mathcal{C}(u(t))}{\mathrm{d}t} \le 0\,, \tag{3.4}$$

where $\Lambda(t)$ is a measurement of the energy per particle that is exchanged with the surrounding. Remember that the considered physical system is the $FePO_4$ electrode. Thus, when Lithium atoms get (de)intercalated, this is an exchange of particles of our system with the surrounding. The term $\Lambda(t)\mathrm{d}\mathcal{C}(u(t))/\mathrm{d}t$ is non-zero only if the system exchanges lithium with the environment.

## 3.2 Deriving the Model Equation

Assume that the system state $u$ changes according to some velocity field $v$ in the domain $\Omega$. We start with the formal assumption that the evolution of $u$ can be described by the

partial differential equation

$$\partial_t u(t,x) + \partial_x[v(t,x)u(t,x)] = 0 \ , \tag{3.5}$$
$$v(0,t) = v(1,t) = 0 \ . \tag{3.6}$$

Inserting this PDE in the derivative of $\mathcal{A}$, see (3.3), and using (3.4) gives

$$\int_a^b \Big(\psi(x) - \Lambda(t)x + \nu^2(1 + \log u(t,x)\Big) \frac{\mathrm{d}u(t,x)}{\mathrm{d}t}\,\mathrm{d}x \le 0 \ .$$
$$\Rightarrow \int_a^b \Big(\psi'(x) - \Lambda(t) + \nu^2\partial_x(\log u(t,x)\Big)\, vu(t,x)\,\mathrm{d}x \le 0 \ ,$$

where we used $v(0) = v(1) = 0$ from which we formally deduce vanishing boundary terms in the partial integration.

In order to guarantee this inequality we demand $v$ to have the form

$$\tau v(t,x) = -\Big(\psi'(x) - \Lambda(t) + \nu^2\partial_x \log u(t,x)\Big)$$

with some $\tau \in \mathbb{R}_+$. Inserting this in (3.5) results in a PDE. Before we actually formulate the PDE we have to mention first, that we want to describe an experiment in which the loading rate is controlled. That means on a time Interval $[0,T]$ a function $\ell : [0,T] \to [0,1]$ is given, which describes the loading state of the system at each time in advance. It imposes therefore an extra condition and serves to determine the yet unknown value for $\Lambda$. Thus we end up with the following Problem.

Let $\tau, \nu, T > 0$, $\Omega = ]0,1[$ and $\psi \in \mathrm{H}^1(\Omega)$. Furthermore assume that $u_0 \in \mathrm{L}^2(\Omega)$ is a given non-negative function with mean value one. We state the PDE

$$\begin{cases} \tau\partial_t u(t,x) = \partial_x\Big(\nu^2\partial_x u(t,x) + \psi'(x)u(t,x) - \Lambda(t)u(t,x)\Big) & \text{for } x \in \Omega,\ t > 0, \\ \nu^2\partial_x u(t,x) + \psi'(x)u(t,x) - \Lambda(t)u(t,x) = 0 & \text{for } x \in \partial\Omega,\ t > 0, \\ \mathcal{C}(u(t)) = \ell(t) & \text{for } t \ge 0, \\ u(0,x) = u_0(x) & \text{for } x \in \Omega. \end{cases} \tag{3.7}$$

The function $\Lambda : [0,T] \to \mathbb{R}$ is unknown and has to be chosen in such a way that the constraint $\mathcal{C}(u(t)) = \ell(t)$ is satisfied. We do not specify the quality of $\ell$. We will derive the needed quality in the later analysis of this model. One immediately sees that $u_0$ must satisfy the compatibility condition $\mathcal{C}(u_0) = \ell(0)$, where $\mathcal{C}$ is defined as in (3.2).

The constraint $\mathcal{C}(u(t)) = \ell(t)$ helps to determine $\Lambda(t)$. We take the derivative of this equality with respect to time, which gives

$$\tau\dot{\ell}(t) = \int_\Omega x\partial_t u(t,x)\,\mathrm{d}x = -\int_\Omega x\partial_x[v(t,x)u(t,x)]\,\mathrm{d}x = \int_\Omega v(t,x)u(t,x)\,\mathrm{d}x \ ,$$
$$= \Lambda(t)\int_\Omega u(t,x)\,\mathrm{d}x - \int_\Omega \psi(x)\ u(t,x)\,\mathrm{d}x - \nu^2\big(u(t,1) - u(t,0)\big) \ .$$

The integral over $u$ is one and so this results in $\Lambda(t) = \mathbb{L}(u(t)) + \tau\dot{\ell}(t)$ where we define for ease of notation in later computation $\mathbb{L} : \mathrm{C}(\Omega) \to \mathbb{R}$ as

$$\mathbb{L}(v) = \int_\Omega \psi'(x)v(x)\,\mathrm{d}x + \nu^2\big(v(1) - v(0)\big)\ . \tag{3.8}$$

We insert this expression for $\Lambda$ in order to resolve the constraint in (3.7) and end up with the following nonlinear Fokker-Planck equation.

$$\begin{cases} \tau\partial_t u(t,x) = \partial_x\Big(\nu^2\partial_x u(t,x) + u(t,x)\Big[\psi'(x) - \mathbb{L}(u(t)) - \tau p(t)\Big]\Big) & \text{for } x \in \Omega,\ t > 0, \\ \nu^2\partial_x u(t,x) + u(t,x)\Big[\psi'(x) - \mathbb{L}(u(t)) - \tau p(t)\Big] = 0 & \text{for } x \in \partial\Omega,\ t > 0, \end{cases} \tag{3.9}$$

where $p = \dot{\ell}$.

Another interesting way to derive (3.7) and (3.9) shall be mentioned here. The general Fokker–Planck equation is a gradient flow with respect to the Wasserstein metric. This has been introduced by Jordan, Kinderlehrer and Otto in [22]. The PDE at hand can be derived as a constrained gradient flow, where the constraint is $\mathcal{C}(u(t)) = \ell(t)$. Since we do not exploit this approach in this work, we refer to [13] for this alternative derivation of the above PDE.

## 3.3 Basic Properties

In this section we deduce the the most obvious properties, for weak solutions $u$ to (3.7) or (3.9) on a time interval $S = [0, T]$.

**Lemma 3.3.1.** *(i) Let $u \in \mathrm{C}^1(S, \mathrm{L}^2(\Omega)) \cap \mathrm{C}(S, \mathrm{H}^1(\Omega))$ be e weak solution to (3.7) or (3.9), then $u$ maintains its mass (mean value), at all times.*
*(ii) Let $\ell$ be differentiable and take $p \equiv \dot{\ell}$ in (3.9). Further assume that the compatibility conditions $\ell(0) = \int_\Omega x u_0(x)\,dx$ and $\int_\Omega u_0(x)\,dx = 1$ are satisfied. A function $u \in \mathrm{C}^1(S, \mathrm{L}^2(\Omega)) \cap \mathrm{C}(S, \mathrm{H}^1(\Omega))$ is a weak solution of (3.7) if and only if it is a weak solution of (3.9). Thus both problem formulations can be treated as equivalent.*

*Proof.* By assumption the solutions have a regularity, so that we can test (3.7) or (3.9) with the function $v(x) = 1$ and integrate by parts. The boundary conditions then guarantee that for solutions $u$ there holds

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_\Omega u(t,x)\,\mathrm{d}x = 0\ , \tag{3.10}$$

which proves (i).

In order to prove (ii) we have to show that a solution of each PDE is a solution of the other. One direction, namely from (3.7) to (3.9), is already done in the Section 3.2, when deriving the PDE (3.9) out of (3.7). The other direction follows when taking $v(x) = x$ as a testfunction in (3.9) and exploiting $\int_\Omega u(x,t) = 1$ for all $t \in S$. Thus, due to the equality $p(t) = \ell'(t)$, solutions of (3.9) satisfy $\ell(t) = \int_\Omega x u(t,x)\,\mathrm{d}x$ and hence (3.7) is satisfied when choosing $\Lambda(t) = \tau\dot{\ell} + \mathbb{L}(u(t))$. For the definition of $\mathbb{L}$ see (3.8). □

The stationary states of the PDE (3.9) are easily specified by setting $\partial_t u \equiv 0$ and solving the resulting spatial ODE. Respecting the condition of mass one, see (3.1), stationary solutions must have the form

$$u_\beta = \frac{1}{c_\beta} \mathrm{e}^{\frac{\beta x - \psi(x)}{\nu^2}} \ , \qquad \text{with} \quad c_\beta = \int_\Omega \mathrm{e}^{\frac{\beta x - \psi(x)}{\nu^2}} \, \mathrm{d}x \ . \tag{3.11}$$

The free parameter $\beta \in \mathbb{R}$ can be determined by prescribing $\mathcal{C}(u_\beta)$.

**Lemma 3.3.2.** *The function $M(\beta) = \mathcal{C}(u_\beta)$ is a bijective mapping from $]-\infty, +\infty[$ to $]0,1[$. Thus for every $\ell \in ]0,1[$ there exists a unique $\beta(\ell)$ such that*

$$U_\ell(x) = \frac{1}{c_{\beta(\ell)}} \mathrm{e}^{\frac{\beta(\ell) x - \psi(x)}{\nu^2}} \qquad \textit{with} \quad c_{\beta(\ell)} = \int_\Omega \mathrm{e}^{\frac{\beta(\ell) x - \psi(x)}{\nu^2}} \, dx \ , \tag{3.12}$$

*is a steady state of (3.9) satisfying $\mathcal{C}(U_\ell) = \ell$.*

*Proof.* First we show that $M$ is strictly increasing on $]-\infty, +\infty[$ and then we will show that it maps onto $]0,1[$. Let us take the derivative of $M$ with respect to $\beta$.

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\beta} M(\beta) &= \frac{\mathrm{d}}{\mathrm{d}\beta} \frac{\int_\Omega x \ \mathrm{e}^{(\beta x - \psi(x))/\nu^2} \, \mathrm{d}x}{\int_\Omega \mathrm{e}^{(\beta x - \psi(x))/\nu^2} \, \mathrm{d}x} \ , \\ &= \frac{\Big( \int_\Omega x^2 \ \mathrm{e}^{(\beta x - \psi(x))/\nu^2} \, \mathrm{d}x \Big) \Big( \int_\Omega \mathrm{e}^{(\beta x - \psi(x))/\nu^2} \, \mathrm{d}x \Big) - \Big( \int_\Omega x \ \mathrm{e}^{(\beta x - \psi(x))/\nu^2} \, \mathrm{d}x \Big)^2}{\Big( \int_\Omega \mathrm{e}^{(\beta x - \psi(x))/\nu^2} \, \mathrm{d}x \Big)^2} \ . \end{aligned}$$

Since $\rho_1(x) := x \mathrm{e}^{\frac{1}{2}(\beta x - \psi(x))}$ and $\rho_2(x) := \mathrm{e}^{\frac{1}{2}(\beta x - \psi(x))}$ are linearly independent functions, the Cauchy-Schwarz inequality says that

$$\Big( \int_\Omega \rho_1 \rho_2 \, \mathrm{d}x \Big)^2 < \Big( \int_\Omega \rho_1^2 \, \mathrm{d}x \Big) \cdot \Big( \int_\Omega \rho_2^2 \, \mathrm{d}x \Big)$$

and thus ensures that $\frac{\mathrm{d}}{\mathrm{d}\beta} \int_\Omega x u_\beta(x) \, \mathrm{d}x > 0$ for each $\beta \in \mathbb{R}$.

What remains to be shown is that the image of $M$ is $]0,1[$. Since $u_\beta$ is a smooth positive function with mean value one, it is obvious that $M(]-\infty, +\infty[) \subset ]0,1[$. Hence we need to verify $M(\beta) \to 0$ as $\beta \to -\infty$ and $M(\beta) \to 1$ as $\beta \to +\infty$ such that the image of $M$ fills out $]0,1[$. We only show the limit for $\beta \to -\infty$ as the other one can be obtained analogous. We claim that for every $\varepsilon > 0$, $\sup\{u_\beta(x) : \ x \in ]\varepsilon, 1[ \ \} \to 0$ as $\beta \to -\infty$, of which $M(\beta) \to 0$ is a simple consequence.

Let us fix $x_0 \in ]0,1]$ and assume that there exist $C_0 > 0$ and a sequence of numbers $\beta_k \to -\infty$ such that $u_{\beta_k}(x_0) \geq C_0$ for all $k$. Since $\psi \in \mathrm{H}^1(\Omega)$ we have $C_1 > 0$ such that $\|\psi\|_{\mathrm{C}} < C_1$ and so $\|\mathrm{e}^{\psi(\cdot)}\|_{\mathrm{C}} \leq C_2$. Then for $\beta < 0$ we can estimate $u_\beta$ on $]0, x_0/2[$ by

$$\frac{|u_\beta(x_0)|}{|u_\beta(x)|} \leq |\mathrm{e}^{\beta(x_0 - x)}| \ |\mathrm{e}^{\psi(x) - \psi(x_0)}| \leq |\mathrm{e}^{\frac{\beta}{2} x_0}| \ C_2^2 \ \xrightarrow{\beta \to -\infty} 0, \qquad \text{for all } x \in \left] 0, \frac{x_0}{2} \right[ .$$

Thus, we can fix $\beta_0 < 0$ such that for all $x \in \left]0, \frac{x_0}{2}\right[$ we get $u_\beta(x) \geq \frac{4}{C_0 x_0} u_\beta(x_0) \geq \frac{4}{x_0}$ whenever $\beta < \beta_0$, which implies that for all sufficiently large $k$

$$1 = \int_\Omega u_{\beta_k} \geq \int_0^{\frac{x_0}{2}} u_{\beta_k} \geq 2\,.$$

This is a contradiction to the construction (3.11), which proves the pointwise convergence to zero of $u_{\beta_k}$ on $]0, 1]$. By the same reasoning we can fix $\beta_1 < 0$ for any $x_0 \in ]0, 1[$, such that for all $x \in ]2x_0, 1]$ we get $u_\beta(x) \leq C_3 u_\beta(x_0)$ whenever $\beta < \beta_1$. This implies the uniform convergence on every subset $]2x_0, 1]$ and thus by the arbitrariness of $x_0$ we get for any $\varepsilon > 0$ that $\sup\{u_\beta(x) : \ x \in ]\varepsilon, 1[\ \} \to 0$ as $\beta \to -\infty$. □

# 4 Analysis of the Nonlinear Fokker–Planck Equation

In this Chapter we will analyse properties of the nonlinear Fokker–Planck equation (3.9) resulting from the modelling of a battery in Chapter 3. In the first section we deal with the existence and uniqueness of solutions for, at least, a short time. This fundamental property allows us to consider the Model as a sane candidate to reflect any physical behaviour. Other properties, like nonnegativity and regularity, are also shown. The nonnegativity is of special interest since we think of the solution as the evolution of a probability density. This interpretation is not adequate if the solution assumes negative values. In the second section the question of long time existence is addressed. It turns out, that a condition, which feels natural from a physical point of view, is a necessary and sufficient condition for solutions to exist up to any considered time. The results presented in this chapter are from a joint work of Dreyer, Huth, Mielke, Rehberg and Winkler, see [13].

For ease of notation we set $\tau = \nu = 1$ throughout this chapter. We can do so without loss of generality, since dividing (3.9) by $\nu^2$ yields the equivalent PDE

$$\begin{aligned}\frac{\tau}{\nu^2}\partial_t u(t,x) &= \partial_x\Big(\partial_x u(t,x) + u(t,x)\big[\frac{\psi'}{\nu^2}(x) - \mathbb{L}_2(u(t)) - \frac{\tau}{\nu^2}p(t)\big]\Big)\,,\\ \mathbb{L}_2(u(t)) &= \frac{\mathbb{L}(v)}{\nu^2} = (v(1)-v(0)) + \int_\Omega \frac{\psi'(x)}{\nu^2}v(x)\,\mathrm{d}x.\end{aligned}$$

This shows, that we can eliminate all occurrences of $\tau$ and $\nu$ on the right hand side by transforming the data to $\tilde{p}(t) = \tau p(t)/\nu^2$ and $\tilde{\psi}'(x) = \psi'(x)/\nu^2$. Another time transformation then easily leads to the disappearance of the factor $\tau/\nu^2$ in front of the time derivative.

## 4.1 Existence and Uniqueness

We formulate our short-time existence result first in a general form of operator differential equations. In this form it can be applied to a great family of nonlinear parabolic problems. The following theorem is similar to results in the book by Lunardi [28]. In the version it is presented here, our result benefits from a dividable dependence of the nonlinear term into parts which depend on the solution in differing regularities. The usual way of dividing would be to separate parts of lowest and a part of highest regularity. We use a more layered separation to benefit from the underlying structure.

First we introduce some notation to formulate a prototype of a nonlinear operator

differential equation. Let $X$ be a Banach space and $S$ be the time interval under consideration. Further on let $A$ be a sectorial operator $A : X \supset D(A) \to X$ with dense domain $D(A)$, see the definition of sectorial operators in the Appendix. We denote by $X_\theta = [X, D(A)]_\theta$ the complex interpolation space to the index $\theta$, and its norm by $\|\cdot\|_\theta$. The possibly nonlinear operator $N : X_\theta \times S \to X$ acts on intermediate spaces $X_\theta$ for some $\theta \in [0, 1[$. The general problem formulation is the following operator differential equation on a time interval $S$,

$$\begin{aligned} \partial_t u(t) &= Au(t) + N\big(u(t), t\big) \ , \qquad\qquad \text{for all } t \in S \\ u(0) &= u_0 \in X \ . \end{aligned} \tag{4.1}$$

Without loss of generality we say that $S$ starts at time zero. If we have a priori knowledge on the nonlinearity of $N$ and also have a better quality of $u_0$, then we can state the following existence and uniqueness result.

**Theorem 4.1.1.** *Let $X$ be a Banach space and $A : D(A) \subset X \to X$ a sectorial operator with dense domain. Define the time interval $S = [0, T]$ and suppose that for $g \in \mathrm{C}(\mathbb{R})$, $0 < \theta_1 < \theta_2 < 1$ and $1 \le p \le (1 - \theta_1)/(\theta_2 - \theta_1)$ the mapping $N : X_{\theta_2} \times S \to X$ satisfies for all $v, w \in X_{\theta_2}$ and all $s \in S$,*

$$\|N(v, s)\|_X \le g(\|v\|_{\theta_1})(1 + \|v\|_{\theta_2}) \ , \tag{4.2}$$

$$\|N(v, s) - N(w, s)\|_X \le g(\|v\|_{\theta_1} + \|w\|_{\theta_1})(\|v - w\|_{\theta_2} + (1 + \|v\|_{\theta_2}^p + \|w\|_{\theta_2}^p)\|v - w\|_{\theta_1}) \ , \tag{4.3}$$

*and $u_0 \in X_{\theta_1}$, then there exists a constant $K$ depending on $A$, $g$, $p$ and $\|u_0\|_{\theta_1}$ such that for $T \le K$, we know that the problem (4.1) has a unique mild solution of quality*

$$u \in Y \ := \ \mathrm{C}(S, X_{\theta_1}) \ \cap \ \mathrm{C}(]0, T[, X_{\theta_2}) \ , \tag{4.4}$$

$$\|u\|_Y := \sup_{t \in S} \|u(t)\|_{\theta_1} + \sup_{t \in S} t^\gamma \|u(t)\|_{\theta_2} \ , \tag{4.5}$$

*for any $\gamma$ satisfying $\theta_2 - \theta_1 < \gamma < (1 - \theta_1)/p$. Furthermore for any $\epsilon > 0$ and $\eta \in [0, 1[$*

$$u \in \mathrm{C}^{1-\eta}\big([\varepsilon, T], X_\eta\big) \ . \tag{4.6}$$

*Proof.* Fix the choice of $\gamma \in ]\theta_2 - \theta_1, (1 - \theta_1)/p[$ and consider the space $Y$ as defined by (4.4) and (4.5). We will show that $F$, defined as

$$Fv(t) := \mathrm{e}^{-tA} u_0 + \int_0^t \mathrm{e}^{-(t-s)A} N(v(s), s)\, \mathrm{d}s, \qquad v \in U, \ t \in [0, T], \tag{4.7}$$

is a contractive selfmapping on the closed subset

$$U := \{v \in Y : \|v\|_Y \le 2 + c_A \|u_0\|_{\theta_1} =: R\} \subset Y. \tag{4.8}$$

Thus $F$ has a unique fixed point in $U$. It is known that fixpoints of $F$ are (mild) solutions to (4.1). The constant $c_A$ appearing in (4.8) depends on the operator $A$ and

will be determined in the sequel.

In order to show that $F$ maps $U$ to itself we derive estimates for $F(v)$ in the $\|\cdot\|_Y$ norm. The first addend in (4.7) can easily be estimated by using known properties for the semigroup $\mathrm{e}^{-tA}$, see Lemma A.3 in the appendix. Thus for any $t \in ]0,T[$

$$\|\mathrm{e}^{-tA}v(0)\|_{\theta_1} \le c_A\|v(0)\|_{\theta_1} \ , \qquad t^\gamma\|\mathrm{e}^{-tA}v(0)\|_{\theta_2} \le c_A t^{\gamma-(\theta_2-\theta_1)}\|v(0)\|_{\theta_1} \ , \tag{4.9}$$

with a constant $c_A \ge 1$ depending on $A$. This is the constant that appears in (4.8). Thus diminishing $T$, depending on $\|v(0)\|_{\theta_1}$ and $c_A$, we can guarantee that $t^\gamma\|\mathrm{e}^{-tA}v(0)\|_{\theta_2} \le 1$.

In order to derive bounds for the second addend in (4.7) we define $C_g := \sup_{s\in[0,2R]} g(s)$ and examine for any $v \in U$ and $t \in [0,T]$,

$$\begin{aligned}\|\int_0^t \mathrm{e}^{-(t-s)A}N\big(v(s),s\big)\,\mathrm{d}s\|_{\theta_1} &\le c_A\int_0^t (t-s)^{-\theta_1}\|N\big(v(s),s\big)\|_X\,\mathrm{d}s \ ,\\ &\le c_A C_g\int_0^t (t-s)^{-\theta_1}(1+\|v(s)\|_{\theta_2})\,\mathrm{d}s \le c\int_0^t (t-s)^{-\theta_1}(1+s^{-\gamma}R)\,\mathrm{d}s,\\ &\le c(1+R)\int_0^t (t-s)^{-\theta_1}s^{-\gamma}\,\mathrm{d}s\end{aligned}$$

and hence by a simple variable transformation from $s$ to $\sigma = s/t$,

$$\begin{aligned}\|\int_0^t \mathrm{e}^{-(t-s)A}N\big(v(s),s\big)\,\mathrm{d}s\|_{\theta_1} &\le c(1+R)t^{1-\theta_1-\gamma}\int_0^1 (1-\sigma)^{-\theta_1}\sigma^{-\gamma}\,\mathrm{d}\sigma,\\ &\le c(1+R)T^{1-\theta_1-\gamma}\int_0^1 (1-\sigma)^{-\theta_1}\sigma^{-\gamma}\,\mathrm{d}\sigma.\end{aligned}$$

We use that $\gamma < (1-\theta_1)/p \le 1-\theta_1$ and get together with (4.9) for $T$ small enough

$$\|F(v)(t)\|_{\theta_1} \le c_A\|v_0\|_{\theta_1} + \frac{1}{2} \ . \tag{4.10}$$

In the same manner we derive estimates in the $\|\cdot\|_{\theta_2}$ norm for any $v \in U$ and $t \in [0,T]$,

$$\begin{aligned}\|\int_0^t \mathrm{e}^{-(t-s)A}N\big(v(s),s\big)\,\mathrm{d}s\|_{\theta_2} &\le c_A\int_0^t (t-s)^{-\theta_2}\|N\big(v(s),s\big)\|_X\,\mathrm{d}s \ ,\\ &\le c_A C_g(1+R)T^{1-\theta_2-\gamma}\int_0^1 (1-\sigma)^{-\theta_1}\sigma^{-\gamma}\,\mathrm{d}\sigma \ ,\end{aligned}$$

which gives by $\theta_2 < 1$ and choosing $T$ small enough together with (4.9),

$$t^\gamma\|F(v)(t)\|_{\theta_2} \le 1 + \frac{1}{2} \ . \tag{4.11}$$

Combining (4.10) and (4.11) results for any $v \in U$ in $\|Fv\|_Y \le R$ , and so $F$ is a selfmapping on $U$.

The steps to show that $F$ is a contraction in the space $Y$ are of similar type. First we estimate for two elements $v, w \in U$ and $t \in [0,T]$ the difference of their images in the

$\|\cdot\|_{\theta_1}$ norm by

$$\begin{aligned}\|Fv(t)-Fw(t)\|_{\theta_1} &\le c\int_0^t (t-s)^{-\theta_1}\Big(\|v-w\|_{\theta_2}+(1+\|v\|_{\theta_2}^p+\|w\|_{\theta_2}^p)\|v-w\|_{\theta_1}\Big)\,\mathrm{d}s\ ,\\ &\le c\|v-w\|_Y\int_0^t\Big((t-s)^{-\theta_1}s^{-\gamma}+(t-s)^{-\theta_1}(1+2R^p s^{-\gamma p})\Big)\,\mathrm{d}s\ ,\\ &\le c\|v-w\|_Y\Big(t^{1-\theta_1-\gamma}\int_0^1(1-\sigma)^{-\theta_1}\sigma^{-\gamma}\,\mathrm{d}\sigma\\ &\qquad + t^{1-\theta_1-\gamma p}(1+2R^p)\int_0^1(1-\sigma)^{-\theta_1}\sigma^{-\gamma p}\,\mathrm{d}s\Big)\ .\end{aligned}$$

The assumptions on $\gamma$ and $p$ allow us to deduce by $t\le T$ that for $T$ small enough there holds

$$\|Fv(t)-Fw(t)\|_{\theta_1}\le\frac{1}{4}\|v-w\|_Y\ . \tag{4.12}$$

Analogously we derive the estimate

$$\begin{aligned}\|Fv(t)-Fw(t)\|_{\theta_2} \le c\|v-w\|_Y\Big(&t^{1-\theta_2-\gamma}\int_0^1(1-\sigma)^{-\theta_1}\sigma^{-\gamma}\,\mathrm{d}\sigma\\ &+t^{1-\theta_2-\gamma p}(1+2R^p)\int_0^1(1-\sigma)^{-\theta_1}\sigma^{-\gamma p}\,\mathrm{d}s\Big)\ .\end{aligned}$$

The assumptions on $\gamma$ and $p$ allow us to deduce once more by $t\le T$ that for $T$ small enough there holds for all $t\in S$

$$t^\gamma\|Fv(t)-Fw(t)\|_{\theta_2}\le\frac{1}{4}\|v-w\|_Y\ . \tag{4.13}$$

In combination of (4.12) and (4.13) we get the contractivity of $F$, by the inequality

$$\|Fv(t)-Fw(t)\|_Y\le\frac{1}{2}\|v-w\|_Y\ .$$

Thus $F$ must have a unique fixpoint in $U$ which proves (4.4) and (4.5).

As a consequence the solution $u$ is for all positive times $t>\varepsilon$ bounded in the $\|\cdot\|_{\theta_2}$ norm and so $N(u(t),t)$ is bounded in $X$ for all these times as well. Inserting this in the right hand side of (4.1) gives a linear operator differential equation with bounded right hand side $N(u(t),t)$ on every time interval $[\varepsilon,T]$ for any $\varepsilon>0$. According to known regularity theory, see [28, Prop.4.2.1], this results in the solution being even Hölder continuous in time in the way as asserted in (4.6). □

We will use this general existence result for the Fokker-Planck equation (3.9). First we define the notion of solutions to the PDE (3.9). By a solution we mean a function

$$u\in \mathrm{C}^1(\,]0,T_0[\,,\mathrm{L}^q(\Omega))\ \cap\ C(\,]0,T_0[\,,\mathrm{W}^{1,q}(\Omega))\ \cap\ C(\,[0,T_0[\,,\mathrm{L}^q(\Omega))\ ,\ \text{where } \Omega:=]0,1[,$$

such that $u$ fulfils the initial condition $u(0) = u_0 \in \mathrm{L}^2(\Omega)$ and that for all $\varphi \in \mathrm{C}^\infty(\bar{\Omega})$ and $t \in ]0, T_0[$ there holds

$$\int_\Omega \partial_t u(t,x)\varphi(x)\,\mathrm{d}x = -\int_\Omega \Big(\partial_x u(t,x) + u(t,x)\Big[\mu(x) - \mathbb{L}(u(t)) - p(t)\Big]\Big)\partial_x\varphi(x)\,\mathrm{d}x. \tag{4.14}$$

The functional $\mathbb{L} : \mathrm{C}(\bar{\Omega}) \to \mathbb{R}$ above is as defined in (3.8). Note that we write $\mu$ for $\psi'$ and $p$ for $\dot{\ell}$. This separates their meaning as coefficients in the PDE from their role as derivatives of another function.

**Theorem 4.1.2.** *Suppose that $p \in \mathrm{C}^\alpha_{loc}([0,\infty[)$, $\mu \in \mathrm{L}^q(\Omega)$, and that $u_0 \in \mathrm{L}^q(\Omega)$ for some $q > 1$ and $\alpha > 1/2$. Then there exists a maximal time $T_0 \in ]0,\infty]$ and a uniquely determined solution $u$ to (4.14). For any $\theta > \left]\frac{1}{q}, 1\right[$ there exists $\delta$, $K > 0$ and $\gamma \in ]0,1[$ depending on $\|u_0\|_{\mathrm{L}^q}$ and the date $\mu$ and $p$ such that*

$$\|u(t)\|_{\mathrm{L}^q} + t^\gamma \|u(t)\|_{\mathrm{W}^{\theta,q}(\Omega)} \le 2 + \|u(0)\|_{\mathrm{L}^q}\ , \qquad \text{for all } t \in ]0,\delta[\ . \tag{4.15}$$

*Moreover we have the following alternative:*

$$\text{Either } T_0 = \infty, \text{ or } \|u(t)\|_{L^q} \to \infty \text{ as } t \nearrow T_0 \quad \text{ for some } q > 1. \tag{4.16}$$

*Proof.* Existence and Uniqueness. First we prove the existence of a solution to a secondary problem. Define $M := \int_0^1 u_0(x)\,\mathrm{d}x$ and $w_0(x) := \int_0^x u_0(z)\,\mathrm{d}z - Mx$ and consider

$$\begin{aligned} \partial_t w(t,x) - \partial_{xx} w(t,x) &= \Big[\mu(x) - p(t) - \mathbb{L}\left(\partial_x w(t) + M\right)\Big](\partial_x w(t,x) + M), \\ w(t,0) &= w(t,1) = 0, \\ w(0,x) &= w_0(x). \end{aligned} \tag{4.17}$$

By regularity results it follows in the sequel that the function $u := \partial_x w + M$ is a solution to the original problem (4.14).

Take an arbitrary but fixed $\beta \in \left]1 + \frac{1}{q}, 2\right]$, then the injection $\mathrm{W}^{\beta,q}_0(\Omega) \hookrightarrow \mathrm{C}^1(\bar{\Omega})$ is compact. Thus remembering the definition of $\mathbb{L}$ in (3.8), a constant $c_{\mathbb{L}}$ exists such that for all $v \in \mathrm{W}^{\beta,q}_0(\Omega)$ there holds

$$|\mathbb{L}(\partial_x v)| \le c_{\mathbb{L}} \|v\|_{\mathrm{W}^{\beta,1}_0} \qquad \text{and} \qquad |\mathbb{L}(\partial_x v + M)| \le c_{\mathbb{L}}(\|v\|_{\mathrm{W}^{\beta,1}_0} + M)\,.$$

Let $A$ denote the Dirichlet Laplacian on $\Omega$. Its domain is $\mathrm{W}^{2,q}_0(\Omega)$, which is dense in $\mathrm{L}^q(\Omega)$. We denote the complex interpolation spaces between them by $\mathrm{W}^{\theta,q}_0(\Omega) = [\mathrm{L}^q, \mathrm{W}^{2,q}_0(\Omega)]_{\frac{\theta}{2}}$. We further define the mapping $N : \mathrm{W}^{\beta,q}_0(\Omega) \times \mathbb{R}_+ \to \mathrm{L}^q(\Omega)$ for $(v,t) \in \mathrm{W}^{\beta,q}_0(\Omega) \times \mathbb{R}_+$ as

$$\Big(Nv\Big)(x,t) := -\Big(\mathbb{L}(\partial_x v + M) - \mu(x) + p(t)\Big) \cdot (\partial_x v(x) + M).$$

In order to apply Theorem 4.1.1, we need to show bounds for $N$ of the form (4.2) and (4.3). By the a priori bounds on $p$, $\mu$ and $\mathbb{L}$ we get for $v, w \in \mathrm{W}_0^{\beta,q}(\Omega)$

$$\begin{aligned}
\|Nv(\cdot,t)\|_{\mathrm{L}^q} &\le c(\|\mu\|_{\mathrm{L}^q} + \|p\|_{\mathrm{L}^\infty})(M + \|v\|_{\mathrm{W}_0^{\beta,q}}) + c(\|v\|_{\mathrm{W}_0^{\beta,1}} + M)\|v\|_{\mathrm{W}_0^{1,1}} \ , \\
\|Nv(\cdot,t) - Nw(\cdot,t)\|_{\mathrm{L}^q} &\le \|(\mu(x) - p(t)) \cdot (\partial_x v - \partial_x w)\|_{\mathrm{L}^q} \\
&\quad + \|\mathbb{L}(\partial_x v(t) + M)(\partial_x v(t,x) + M) \\
&\qquad - \mathbb{L}(\partial_x w(t) + M)(\partial_x w(t,x) + M)\|_{\mathrm{L}^q} \ , \\
&\le c(\|\mu\|_{\mathrm{L}^q} + \|p\|_{\mathrm{L}^\infty})(\|v - w\|_{\mathrm{W}_0^{\beta,q}}) \\
&\quad + c(\|w\|_{\mathrm{W}_0^{\beta,q}} + M)\|v - w\|_{\mathrm{W}_0^{1,q}} \\
&\quad + \|v - w\|_{\mathrm{W}_0^{\beta,q}} c(\|w\|_{\mathrm{W}_0^{1,q}} + M) \ .
\end{aligned}$$

Consequently, if we take $p = 1$, $\theta = \frac{1}{2}$ and $\theta_2 = \frac{\beta}{2}$ meaning $X = \mathrm{L}^q(\Omega)$, $X_{\theta_1} := \mathrm{W}_0^{1,q}(\Omega)$ and $X_{\theta_2} := \mathrm{W}_0^{\beta,q}(\Omega)$ we can satisfy the assumptions of Theorem 4.1.1. This gives unique existence of solutions in $\mathrm{C}([0,T], \mathrm{W}_0^{1,q}(\Omega)) \ \cap \ C_\gamma(]0,T], \mathrm{W}_0^{\beta,q}(\Omega))$ for any $\gamma$ satisfying $(\beta - 1)/2 < \gamma < \frac{1}{2}$. Hence (4.15) is a consequence of Theorem 4.1.1 when going back to $u$ by $u = \partial_x w + M$.

The length of the existence time $T$ depends on bounds for $\|\mu\|_{L^q}$, $M$, $\|p\|_{L^\infty}$ and the $\mathrm{W}_0^{1,q}(\Omega)$ norm of the initial value $w_0$. Thus by successively repeating the above local result, one reaches a maximal existence time $T_0$. Provided that we have a uniform bound on $|p(t)|$, the maximal existence time $T_0$ is then either $\infty$ or if $T_0 < \infty$ then there must hold for every $q > 1$, that $\|w(t)\|_{\mathrm{W}_0^{1,q}} \to \infty$ or equivalently $\|u(t)\|_{\mathrm{L}^q} \to \infty$ as $t \to T_0$.

Regularity. The application of Theorem 4.1.1 gives that for any $\eta \in [0, 2[$, the solution $w$ to the above PDE is of quality

$$w \in \mathrm{C}^{1-\frac{\eta}{2}}\big([\varepsilon, T], \mathrm{W}_0^{\eta,q}(\Omega)\big).$$

Now we use the assumption that $p$ is Hölder continuous with the Hölder exponent $\alpha > 0$. The right hand side $N(w(t), t)$ is then Hölder continuous in time on any interval $[\varepsilon, T]$. Thus again using known regularity theory, see [28, Prop.4.3.4], we have that $w$ is a classical solution to (4.17). By this we can iteratively improve the Hölder continuity of $w$ to get by the assumption $\alpha > 1/2$, that there exists a small $\tilde{\alpha} > 0$ such that

$$w \in \mathrm{C}^{1+\alpha}\big([\varepsilon, T], \mathrm{L}^q(\Omega)\big) \ \cap \ \mathrm{C}^{\alpha}\big([\varepsilon, T], \mathrm{W}_0^{2,q}(\Omega)\big) \subset \mathrm{C}^{1+\tilde{\alpha}}\big([\varepsilon, T], \mathrm{W}_0^{1,q}(\Omega)\big).$$

Finally, we have that the time derivative $\partial_t w$ is spatial weakly differentiable for positive times $t$ and vanishes at the boundary of $\Omega$. Thus, we see that $u := \partial_x w - M$ is a solution to (3.9) in the sense of (4.14), since for any $\varphi \in \mathrm{C}^\infty(\bar{\Omega})$ we have

$$\int_\Omega \partial_t u(t,x)\varphi(x)\,\mathrm{d}x = -\int_\Omega \partial_t w(t,x)\partial_x \varphi(x)\,\mathrm{d}x,$$

and hence by partial integration we derive

$$\begin{aligned}\int_\Omega \partial_t u(t,x)\varphi(x)\,\mathrm{d}x &= -\int_\Omega \big(\partial_{xx} w(t,x)\\ &\qquad + (\partial_x w(t,x)+M)\Big[\mu(x)-\mathbb{L}(\partial_x w(t)+M)-p(t)\Big]\big)\partial_x\varphi(x)\,\mathrm{d}x,\\ &= -\int_\Omega \big(\partial_x u(t,x)+u\Big[\mu(x)-\mathbb{L}(u)-p(t)\Big]\big)\partial_x\varphi(x)\,\mathrm{d}x.\end{aligned}$$

□

**Remark 4.1.3.** *The assumption $\dot{\ell} = p \in \mathrm{C}^\alpha_{loc}([0,\infty[)$ can be weakened. If $p$ has this regularity only piecewise on intervals $[t_j, t_{j+1}]$ then one can apply the above theory successively on each interval.*

The above existence result is valid for initial data $u_0 \in \mathrm{L}^q$. This is necessary for the argumentation to carry this short time existence result to a long time existence result which relies on deriving bounds in $\mathrm{L}^q$. The price of this is that the solution $u(t)$ is not bounded in $\mathrm{W}^{1,q}(\Omega)$, as it might leave $\mathrm{W}^{1,q}(\Omega)$ for $t \to 0$. As stated in the next corollary a better quality of the initial datum would also give us boundedness in $\mathrm{W}^{1,q}(\Omega)$ .

**Corollary 4.1.4.** *If the assumptions of Theorem 4.1.2 are satisfied and furthermore $u_0 \in \mathrm{W}^{1,q}(\Omega)$, then we have $u \in \mathrm{C}([0,T], \mathrm{W}^{1,q}(\Omega))$.*

*Proof.* In the same way as before, we apply the Theorem 4.1.1. As before we first look for solutions of (4.17) and choose a $\beta > 1 + 1/q$ such that there holds the compact embedding $\mathrm{W}^{\beta,q}_0(\Omega) \hookrightarrow \mathrm{C}^1(\bar{\Omega})$. But in contrast to the approach in Theorem 4.1.2, we set $\theta_1 = \beta/2$ and $\theta_2 = \beta/2 + \varepsilon < 1$ for some arbitrary small $\varepsilon > 0$.

Since the choices for $\theta_1$ and $\theta_2$ are bigger than in the previous theorem, all the needed bounds in the norms $\|\cdot\|_{\theta_\ell}$ are a trivial consequence of the already achieved bounds in the proof of Theorem 4.1.2. Together with the boundedness of the initial value $\|u_0\|_{\mathrm{W}^{1,q}(\Omega)}$ we derive, for $w$ the solution to (4.17) that

$$w \in \mathrm{C}\big(S, \mathrm{W}^{\beta,q}(\Omega)\big) \;\cap\; \mathrm{C}_\gamma\big(S, \mathrm{W}^{\beta+\varepsilon,q}(\Omega)\big)\ ,$$

and thus $N(w(t), s)$ is bounded in $\mathrm{L}^q(\Omega)$ for all times $t \in S$. Then classical regularity theory gives that $w$ must have the quality

$$w \in \mathrm{C}\big(S, \mathrm{W}^{2,q}(\Omega)\big) \;\cap\; \mathrm{C}^1\big(S, \mathrm{L}^q(\Omega)\big)\ ,$$

and consequently for $u = \partial_x w - M$ there holds

$$u \in \mathrm{C}\big(S, \mathrm{W}^{1,q}(\Omega)\big)\ .$$

□

**Lemma 4.1.5.** *Let the assumptions of Theorem 4.1.2 with $q = 2$ hold. If the initial value $u_0 \in \mathrm{L}^2(\Omega)$ is non-negative, then the solution $u$ to (4.14) is also nonnegative on $\Omega \times [0,T]$.*

*Proof.* We first show the non-negativity if $\mu$ is a bounded function. Since $q = 2$ we can apply the negative part $u^-(t)$ as a test function in (4.14) to get

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u^-(t)\|_{L^2}^2 &\leq -\|\partial_x u^-(t)\|_{L^2}^2 + \int_\Omega \partial_x u^-(t,x)u^-(t,x)(\mu - \mathbb{L}(u(t)) - p(t))\,\mathrm{d}x,\\ &\leq \frac{1}{4}(\|\mu\|_{L^\infty} + c_{\mathbb{L}}Rt^{-\gamma} + \|p\|_{L^\infty})^2\|u^-(t)\|_{L^2}^2,\end{aligned}$$

where we used the continuity of $\mathbb{L}$ and the boundedness of the solution stated in (4.15). Using Gronwall's lemma and $\|u_0^-\|_{L^2} = 0$, we deduce $\|u^-(t)\|_{L^2} = 0$ for all $t \in [0,T]$. If $\mu$ is unbounded, we define for any $k > 0$ the cut-off function $\mu_k := \min(k, \max(-k, \mu))$. By $u_k$ we call the solution of (4.14), if $\mu$ is therein replaced by $\mu_k$. We consider the functions $w_k(t,x) := \int_0^x u_k(y,t)\,\mathrm{d}y - Mx$, being the solution to (4.17) when $\mu$ is replaced by $\mu_k$. Let the space $Y$ and the closed convex subset $U \subset Y$ be as in (4.8), with $X = \mathrm{L}^q(\Omega)$, $X_{\theta_1} = \mathrm{W}^{1,q}(\Omega)$ and $X_{\theta_1} = \mathrm{W}^{\beta,q}(\Omega)$ where $\beta$ is as defined in the proof of the Theorem 4.1.2. Since $\|\mu_k\|_{L^2} \leq \|\mu\|_{L^2}$, one can find a common $T > 0$ such that each of the mappings

$$F_k : U \to U, \tag{4.18}$$

$$F_k v(t) := \mathrm{e}^{-tA}w_0 + \int_0^t \mathrm{e}^{-(t-s)A}\left[\mu_k - \mathbb{L}(\partial_x v(s) + M) - p(s)\right](\partial_x v(s) + M)\,\mathrm{d}s, \tag{4.19}$$

is a contraction on $U$ with contraction constant $\frac{1}{2}$. This is true, because all necessary estimates in the proof of Theorem 4.1.2 can be repeated with the uniform bound $\|\mu_k\|_{L^2} \leq \|\mu\|_{L^2}$. The fixed point of $F_k$ is then $w_k$ and the corresponding fixed point for the mapping $F$ having the original $\mu$, is $w$. One easily calculates

$$\|w - w_k\|_X \leq \|Fw - F_k w_k\|_X \leq \|F_k w - F_k w_k\|_X + \|Fw - F_k w\|_X \tag{4.20}$$

$$\leq \frac{1}{2}\|w - w_k\|_X + \|Fw - F_k w\|_X, \tag{4.21}$$

what leads to $\|w - w_k\|_X \leq 2\|Fw - F_k w\|_X$. Let us show that $\|Fw - F_k w\|_X$ vanishes as $k \to \infty$. We have

$$\begin{aligned}\|Fw(t) - F_k w(t)\|_{H^1} &\leq \|\int_0^t \mathrm{e}^{-(t-s)A}(\partial_x w(s) + M)(\mu - \mu_k)\,\mathrm{d}s\|_{H^1},\\ &\leq \|\mu - \mu_k\|_{L^2} t^{\frac{1}{2}-\gamma}\int_0^1 (1-\sigma)^{-\frac{1}{2}}\sigma^{-\gamma}R_M\,\mathrm{d}\sigma.\end{aligned}$$

Since $\mu_k \to \mu$ in $\mathrm{L}^2(\Omega)$, we get, uniformly in $t$, $\|Fw(t) - F_k w(t)\|_{H^1} \to 0$ as $k \to \infty$. Similarly we get $t^\gamma\|Fw(t) - F_k w(t)\|_{H^\beta} \to 0$ uniformly for all $t \in [0,T]$. This gives for all $t \in [0,T]$ that $w_k(t) \to w(t)$ in $\mathrm{H}^1(\Omega)$, and thus $u_k(t) \to u(t)$ in $\mathrm{L}^2(\Omega)$. Hence, $u(t)$ must also be a nonnegative function because each $u_k(t)$ is, which is shown in the first part of this proof. □

In the model derived in [11], $\mu$ is not only in $\mathrm{L}^2(\Omega)$, but inside of $\Omega$ it is a smooth

function. This helps us to deduce strict positivity of the solution for positive times.

**Lemma 4.1.6.** *Let the assumptions of Theorem 4.1.2 hold. Furthermore assume, that for all $\varepsilon > 0$ we have $\mu \in \mathrm{C}^1([\varepsilon, 1-\varepsilon])$ and $0 \neq u_0 \in \mathrm{L}^2(\Omega)$ is non-negative. The solution is then strictly positive inside $\Omega$ for all positive times.*

*Proof.* Let $u$ be the solution to problem (4.14). We define $\Omega_\varepsilon := ]\varepsilon, 1-\varepsilon[$ and

$$\phi(t,x) := \mu(x) - \mathbb{L}(u(t)) - p(t).$$

Consider the function $w_\varepsilon(t,x) := u(t,x)\mathrm{e}^{st}$ with $s_\varepsilon := -\sup_{x\in\Omega_\varepsilon} |\partial_x \mu(x)| \geq 0$. This function then solves inside $\Omega_\varepsilon \times ]\varepsilon, T]$

$$\partial_t w_\varepsilon(t,x) - \partial_{xx} w_\varepsilon(t,x) - \partial_x w_\varepsilon(t,x)\phi(t,x) = u(t,x)\mathrm{e}^{st}(\partial_x \mu(x) + s_\varepsilon). \tag{4.22}$$

The coefficients and initial values and the boundary values are spatially continuous and in time even Hölder continuous. This allows us to apply classical parabolic theory. We know from Theorem 4.1.2 that the initial and boundary values to this PDE are non-negative. Due to conservation of mass, the initial function $w_{\varepsilon 0}(x) := u(\varepsilon, x)\mathrm{e}^{\varepsilon s}$ is positive somewhere inside $\Omega_\varepsilon$ for $\varepsilon$ small enough. Even the right hand side of the PDE (4.22) is non-negative. Hence, using classical maximum principles, see for example [16, Chapter 7.1 Theorem 9], we get $w > 0$ in $\Omega_\varepsilon \times ]\varepsilon, T]$ and thus $u$ is also strictly positive. This means by the arbitrariness of $\varepsilon$, that $u$ is positive everywhere inside $\Omega$ for all positive times $t$. □

## 4.2 Longtime Existence and Blow Up

In this Section we will investigate the maximal existence time of solutions to the PDE (3.9). Due to the nonlinear nature of the PDE, blow up can occur. As we will see, it is easy to construct data, such that the solution must cease to exist in finite time. A question is, which characteristics of the data allow us to decide a priori if the solution exists for all times and can we find a necessary and sufficient criterium. In fact the sufficient and necessary condition is somewhat obvious from the physical point of view. It turns out to be the need of the loading state $\ell$ to stay in the interval $]0,1[$. The actual proof that the natural condition is the exact criterion to decide if the solution exists for all times or not, is quiet involved as can be seen in the following.

The following natural assumption is necessary and sufficient for the solution to exist on the whole (possibly unbounded) time interval $S = [0, T_0]$.

$$\ell \in \mathrm{W}^{1,\infty}(S)\ , \qquad \ell(0) = \int_\Omega x u_0(x)\,\mathrm{d}x\ , \tag{4.23}$$

$$\exists\ \delta > 0\ :\ \forall\ t \in S : \ell(t) \in [\delta, 1-\delta]\ , \qquad \|\dot{\ell}\|_{\mathrm{L}^\infty} \leq \frac{1}{\delta}\ . \tag{4.24}$$

Next to this assumption we do of course need that the assumptions for the existence in Theorem 4.1.2 are satisfied. That is $p = \dot{\ell}$ is piecewise Hölder continuous in time with

Hölder exponent $\alpha > 1/2$ and $\psi \in \mathrm{W}^{1,q}(\Omega)$.

The necessity of (4.23) and (4.24) can be seen quite easily.

**Lemma 4.2.1.** *Assume that on a compact time interval $S$ the continuous function $u : S \to \mathrm{W}^{1,q}(\Omega)$, $q > 1$, is a solution to (3.7) with $u_0 \geq 0$, $\int_\Omega u_0(x) = 1$ and $\ell(t) \in \mathrm{W}^{1,\infty}(S)$.*

*Then there exists a $\delta > 0$, such that $\ell(t) \in [\delta, 1-\delta]$ for all $t \in S$.*

*Proof.* By the assumption $u_0 \geq 0$ and $\int_\Omega u_0(x) = 1$ for $u_0 \in \mathrm{L}^q(\Omega)$, it is clear that $0 < \ell(0) = \int_\Omega x u(t,x)\,\mathrm{d}x < 1$. Due to the continuity of the solution in $\mathrm{W}^{1,q}(\Omega)$, we also have that $\ell(t) = \mathcal{C}(u(t))$ is continuous in time. Assume that there is a time $t_*$ such that $\int_\Omega x u(t,x)\,\mathrm{d}x = 1$ for the first time. Then there must be a sequence $t_j < t_*$ with $t_j \to t_*$ such that $\mathcal{C}(u(t_j)) \to 1$ and $\mathcal{C}(u(t_j)) \in ]0,1[$.

As known from Lemma 4.1.5, all functions $u(t_j) \in \mathrm{W}^{1,q}(\Omega)$ are nonnegative and so we deduce from $\int_\Omega x u(t_j,x)\,\mathrm{d}x \to 1$ that $u(t_j)$ evolves to a delta distribution concentrated at $x = 1$ and thus $\|u(t_j)\|_{\mathrm{L}^q} \to \infty$. This contradicts the continuity of the trajectory $u$ in the space $\mathrm{W}^{1,q}(\Omega)$. Hence such a time $t_*$ can not exist. The same reasoning works *mutatis mutandis* assuming there is a time $t_*$ such that $\ell(t_*) = 0$. □

In order to show that (4.23) and (4.24) are not only necessary but also sufficient conditions for solutions to exist, we examine the energy

$$\mathcal{A}(v) = \int_\Omega v(x)\big(\ln v(x) + \psi(x)\big)\,\mathrm{d}x\ , \qquad 0 \leq v \in \mathrm{W}^{1,q}(\Omega)\ . \tag{4.25}$$

This energy stems from the modelling process, see (3.3). Once we have established a bound for this energy we will then derive a uniform bound for the solution in the $\mathrm{L}^q(\Omega)$ norm, which according to Theorem 4.1.2 results in the existence of solutions for all times.

We define the dissipation $\mathcal{D}$. This is the negative rate of change of $\mathcal{A}$ along solution trajectories $u : S \to \mathrm{W}^{1,q}(\Omega)$, and can be evaluated to

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{A}(u(t)) = -\mathcal{D}(u(t), \dot\ell(t)) \tag{4.26}$$

$$\mathcal{D}(v,\lambda) = \int_\Omega \frac{\big(\partial_x v(x) + \psi'(x)v(x)\big)^2}{v}\,\mathrm{d}x - \mathbb{L}(v)^2 - \lambda\mathbb{L}(v)\ . \tag{4.27}$$

The main ingredients are the two estimates for $\mathcal{D}$ in the following theorem and proposition. The proofs rely on the exact form of $\mathbb{L}$ and the shape of the stationary states.

**Theorem 4.2.2.** *Assume $\psi \in \mathrm{W}^{1,2}([0,1])$. Then, for each $\delta \in ]0,1/2[$ there exists a constant $C_\delta^\psi \geq 0$ such that for all $\ell \in [\delta, 1-\delta]$ and all $\lambda \in [-1/\delta, 1/\delta]$ the following estimate holds:*

$$\mathcal{D}(u,\lambda) \geq -C_\delta^\psi|\lambda| \quad \text{for all } u \in \mathrm{H}^1(\Omega) \text{ with } \int_\Omega u(x)\,dx = 1 \text{ and } \mathcal{C}(u) = \ell. \tag{4.28}$$

*Proof.* The proof consists of two main ingredients. First, we replace $u$ by $v = \sqrt{u}$. Secondly, we will decompose $v$ into $V_\alpha + \eta$, where $V_\alpha$ is the square root of a stationary

solution as described in (3.11). We can easily prove (4.28) for stationary states, as they make the first two terms of $\mathcal{D}$ cancel each other. We then control the error added by the occurrence of $\eta$.

The prescription of the zeroth and first moments of $u$ translates for $v = \sqrt{u}$ to

$$\int_\Omega v(x)^2 \,\mathrm{d}x = 1 \quad \text{and} \quad \int_\Omega x\, v(x)^2 \,\mathrm{d}x = \ell. \tag{4.29}$$

For $v \in \mathcal{V}(\ell) := \{\, v \in \mathrm{H}^1(\Omega) \,:\, v \geq 0, \text{ and (4.29) holds } \}$ we define

$$w := 2\partial_x v + \psi' v, \quad \gamma := \|w\|_{L^2}, \quad \text{and } \rho := \int_\Omega vw \,\mathrm{d}x \;,$$
$$D(v,\lambda) := \mathcal{D}(v^2,\lambda) = \int_\Omega w^2 \,\mathrm{d}x - \Big(\int vw \,\mathrm{d}x\Big)^2 - \lambda \int_\Omega vw \,\mathrm{d}x = \gamma^2 - \rho^2 - \lambda\rho \;.$$

The case $\rho = 0$ is trivial, because it gives $D(v,\lambda) \geq 0$. Hence, we assume $\rho > 0$ from now on. We define for $\alpha \in \mathbb{R}$ the function $V_\alpha$ as the square root of a stationary state $V_\alpha(x) = c_\alpha \mathrm{e}^{\frac{\alpha x - \psi(x)}{2}}$ with $c_\alpha$ such that $\|V_\alpha\|_{\mathrm{L}^2} = 1$. Then we consider the decomposition

$$v = V_{\gamma^2/\rho} + \eta \;.$$

In order to examine the error $\eta$, we make several steps. Remember, that we assume $\rho > 0$ and hence $\gamma > 0$. We first decompose $v$ in the form

$$v = \frac{\rho}{\gamma^2}\, w + \xi \quad \text{with } \int_\Omega \xi\, w \,\mathrm{d}x = 0,$$

which is a simple orthogonal projection. Hence, we find

$$1 = \|v\|_{L^2}^2 = \frac{\rho^2}{\gamma^2} + \|\xi\|_{L^2}^2 \quad \Rightarrow \quad \|\xi\|_{L^2}^2 = 1 - \frac{\rho^2}{\gamma^2}.$$

Recalling the definition of $w$ in terms of $v$ leads to $2\partial_x v + \psi' v = w = \frac{\gamma^2}{\rho}(v - \xi)$. Solving this ODE with $\|v\|_{L^2} = 1$ gives the formulae

$$v = \beta V_{\gamma^2/\rho} + \mathcal{K}_{\gamma^2/\rho}\xi \;, \qquad \text{where} \quad \mathcal{K}_\alpha \xi(x) = \int_0^1 K_\alpha(x,y)\xi(y)\,\mathrm{d}y \;,$$
$$\eta = (1-\beta) V_{\gamma^2/\rho} + \mathcal{K}_{\gamma^2/\rho}\xi \;.$$

The kernel $K_\alpha$ is defined via

$$K_\alpha(x,y) = \begin{cases} \frac{\alpha V_\alpha(x)}{2V_\alpha(y)} & \text{for } \alpha > 0 \text{ and } 0 < x < y < 1, \\ -\frac{\alpha V_\alpha(x)}{2V_\alpha(y)} & \text{for } \alpha < 0 \text{ and } 0 < y < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

We know that $\psi \in \mathrm{W}^{1,1}(\Omega)$ and hence $\psi \in \mathrm{C}(\bar{\Omega})$. Thus, we can estimate the kernel by

$$0 \le K_\alpha(x,y) \le \frac{C_K^\psi}{2}|\alpha| \mathrm{e}^{-|\alpha||x-y|/2} \text{ for } \alpha \neq 0 \text{ and } x,y \in [0,1],$$

where $C_K^\psi$ depends only on $\psi$ but not on $\alpha$. Using this, we can estimate $\widehat{\xi} := \mathcal{K}_{\gamma^2/\rho}\xi$ via $|\widehat{\xi}(x)| \le C_K^\psi \frac{\alpha}{2} \int_0^1 \mathrm{e}^{-|\alpha||x-y|/2}|\xi(y)|\,\mathrm{d}y$. Then using Young's inequality for convolutions $\widehat{\xi} = \phi * \xi$ in the form $\|\widehat{\xi}\|_{L^2(\mathbb{R})} = \|\phi\|_{L^1(\mathbb{R})}\|\xi\|_{L^2(\mathbb{R})}$, we get the uniform estimate

$$\|\mathcal{K}_\alpha\|_{\mathrm{Lin}(L^2(\Omega),L^2(\Omega))} \le C_K^\psi \text{ for all } \alpha \neq 0.$$

It is now essential to estimate $\eta$ in terms of $\rho/\gamma$. We do this with the help of $\widehat{\xi} = K_{\gamma^2/\rho}\xi$, which satisfies $\|\widehat{\xi}\|_{L^2} \le C_K^\psi(1-\rho^2/\gamma^2)^{1/2}$. So for $\rho/\gamma$ close to one, we can assume $\|\widehat{\xi}\|_{L^2} \le 1$ and hence use the relation for $\beta$,

$$1 \ge \|\widehat{\xi}\|_{L^2}^2 = \|v - \beta V_{\gamma^2/\rho}\|_{L_2}^2 = 1 - 2\beta \int_\Omega v V_{\gamma^2/\rho}\,\mathrm{d}x + \beta^2.$$

We know that $\int_\Omega v V_{\gamma^2/\rho}\,\mathrm{d}x > 0$ and thereby conclude $\beta \ge 0$. Hence,

$$\begin{aligned} |1-\beta| \le |1-\beta^2| &= |\|v\|_{L_2}^2 - \|\beta V_{\gamma^2/\rho}\|_{L_2}^2| = |\int_\Omega (v-\beta V_{\gamma^2/\rho})(v+\beta V_{\gamma^2/\rho})\,\mathrm{d}x| \\ &\le |\int_\Omega \widehat{\xi}(2v+\widehat{\xi})\,\mathrm{d}x| \le (2+C_K^\psi)\|\widehat{\xi}\|_{L_2}. \end{aligned}$$

Combing this with the definition of $\eta$ we find

$$\|\eta\|_{L^2} \le |\beta-1| + \|\widehat{\xi}\|_{L^2} \le (3+C_K^\psi)C_K^\psi(1-\rho^2/\gamma^2)^{1/2} \text{ if } \|\widehat{\xi}\|_{L^2} \le 1. \tag{4.30}$$

Now we are ready to estimate $D(v,\lambda)$ from below on the admissible set $\mathcal{V}(\ell)$. We choose $\sigma_\delta \in ]0,1[$, such that

$$(3+C_K^\psi)C_K^\psi(1-\sigma_\delta^2)^{1/2} \le \delta/6 < 1.$$

and distinguish two cases $\rho^2 \le \gamma^2\sigma_\delta^2$ and $\rho^2/\gamma^2 \in [\sigma_\delta^2, 1]$.

*Case I, $|\rho| \le \gamma\sigma_\delta$:* We easily find

$$D(v,\lambda) = \gamma^2 - \rho^2 - \lambda\rho \ge \gamma^2 - \gamma^2\sigma_\delta^2 - |\lambda|\gamma\sigma_\delta \ge -\frac{\lambda^2\sigma_\delta^2}{4(1-\sigma_\delta^2)} \ge -\frac{\sigma_\delta^2}{4\delta(1-\sigma_\delta^2)}\,|\lambda|\,,$$

where $\delta$ is from the statement of the theorem such that $|\lambda| \le 1/\delta$.

*Case II, $\rho^2/\gamma^2 \in [\sigma_\delta^2, 1]$:* Recalling $\|\widehat{\xi}\|_{L^2} \le C_K^\psi(1-\rho^2/\gamma^2)^{1/2}$, we have $\|\widehat{\xi}\|_{L^2} \le 1$ and can use estimate (4.30) for $\eta$, namely $\|\eta\|_{L^2} \le \delta/6$. Since $v = V_{\gamma^2/\rho} + \eta$ lies in $V(\ell)$, we

obtain

$$\Big|\ell - \int_\Omega x V^2_{\rho^2/\gamma}(x)\,\mathrm{d}x\Big| = \Big|\int_\Omega x(v^2(x) - V^2_{\gamma^2/\rho}(x))\,\mathrm{d}x\Big| \le \int_\Omega \big|2x\eta V_{\gamma^2/\rho} + x\eta^2(x)\big|\,\mathrm{d}x$$
$$\le 2\|\eta\|_{L^2} + \|\eta\|^2_{L^2} \le 3\|\eta\|_{L^2} \le \delta/2.$$

We consider the function $m(\alpha) := \int_\Omega x V_\alpha(x)^2\,\mathrm{d}x\,\mathrm{d}x = \int_\Omega x u_\alpha(x)\,\mathrm{d}x$. According to Lemma 3.3.2, $m : \mathbb{R} \to ]0,1[$ is bijective and strictly increasing. Thus, for each $\delta \in ]0,1/2[$ there is a constant $a_\delta$, such that $m(\alpha) \in [\delta/2, 1-\delta/2]$ implies $\alpha \in [-a_\delta, a_\delta]$.

Using the assumption $\ell \in [\delta, 1-\delta]$ we have shown that the decomposition $v = V_{\gamma^2\rho} + \eta$ implies $m(\gamma^2/\rho) \in [\delta/2, 1-\delta/2]$. Thus, we conclude the estimate $a_\delta \ge |\gamma^2/\rho| \ge |\gamma|$, because $0 < |\rho| \le \gamma$. We insert this into the formula for $D$ and obtain the lower bound

$$D(v,\lambda) = \gamma^2 - \rho^2 - \lambda\rho \ge -a_\delta|\lambda|.$$

Combining the two cases, we have established the desired estimate (4.28) with $C^\psi_\delta = \max\{a_\delta, \sigma^2_\delta/(4\delta(1-\sigma^2_\delta))\}$. □

We further improve this and show that the dissipation can be bounded from below in terms of the $\mathrm{L}^2$ norm of $u$ and even in terms of the Energy $\mathcal{A}$ itself.

**Proposition 4.2.3.** *Assume $\psi \in \mathrm{H}^1(\Omega)$. Then, for each $\kappa > 0$ and each $\delta \in ]0,1/2[$ there exists a constant $K^\psi_{\kappa,\delta}$ such that for all $\ell \in [\delta, 1-\delta]$ and all $\lambda \in [-1/\delta, 1/\delta]$ the following estimate holds:*

$$\mathcal{D}(u,\lambda) \ge \kappa\|\partial_x u\|_{L^2} - K^\psi_{\kappa,\delta} \quad \text{and} \quad \mathcal{D}(u,\lambda) \ge \kappa\mathcal{A}(u) - K^\psi_{\kappa,\delta} \tag{4.31}$$

*for all $u \in \mathrm{H}^1(\Omega)$ with $u \ge 0$, $\int_\Omega u\,dx = 1$, and $\mathcal{C}(u) = \ell$.*

*Proof.* We proceed exactly as in the proof of Theorem 4.2.2 and use the same notations.
*Step 1:* We first estimate

$$D_\kappa(v,\lambda) = D(v,\lambda) - \kappa\|\partial_x v\|^{3/2}_{L^2}.$$

Because of $\gamma = \|2\partial_x v + \psi' v\|_{L^2}$ and $\|v\|_{L^2} = 1$ we can deduce with the help of the Gagliardo-Nirenberg inequality (3), $\|\partial_x v\|_{L^2} \le \gamma + 1 + \frac{1}{2}\|\psi'\|^2_{L^2}$ and find

$$D_\kappa(v,\lambda) \ge \gamma^2 - \rho^2 - \lambda\rho - \kappa\gamma^{3/2} - C\ ,$$

where $C$ depends on $\psi$ and $\kappa$. This can be estimated from below via the two cases as before.

*Case I, $|\rho| \le \gamma\sigma_\delta$:* We obtain

$$D_\kappa(v,\lambda) \ge (1-\sigma^2_\delta)\gamma^2 - \tfrac{1}{\delta}\sigma_\delta\gamma - \kappa\gamma^{3/2} - C,$$

which is certainly bounded from below by a constant depending only on $\kappa$ and $\sigma_\delta$.

*Case II, $\rho^2/\gamma^2 \in [\sigma_\delta^2, 1]$:* As in the previous proof we find $|\rho| \le \gamma \le a_\delta$, giving

$$D_\kappa(v,\lambda) \ge \gamma^2 - \rho^2 - \tfrac{1}{\delta}|\rho| - \kappa\gamma^{3/2} - C$$

is trivially bounded from below.

Combining the two cases gives a constant $k^\psi_{\kappa,\delta}$ such that $D_\kappa(v,\lambda) \ge k^\psi_{\kappa,\delta}$.

*Step 2:* We now need to undo the substitution $u = v^2$ in $\mathcal{D}(u,\lambda) = D(\sqrt{u},\lambda)$. With $\partial_x u = 2v\partial_x v$ we find

$$\|\partial_x u\|_{L^2}^2 = 4\|v\partial_x v\|_{L^2}^2 = 4\|v\|_{L^\infty}^2\|\partial_x v\|_{L^2}^2 \le C(1+\|\partial_x v\|_{L^2}^3),$$

where we have used $\|v\|_{L^\infty}^2 \le C\|v\|_{L^2}(\|v\|_{L^2}+\|\partial_x v\|_{L^2}) = C(1+\|\partial_x v\|_{L^2})$, see Lemma A.2. Using $v = \sqrt{u}$, we deduce

$$\mathcal{D}(u,\lambda) - \kappa\|\partial_x u\|_{L^2} \ge D(v,\lambda) - c_1\kappa\|\partial_x v\|_{L^2}^{3/2} - c_2 \;=\; D_{c_1\kappa}(v,\lambda) - c_2 \;\ge\; k^\psi_{c_1\kappa,\delta} - c_2.$$

Thus, the first estimate in (4.31) is established with $K^\psi_{\kappa,\delta} := k^\psi_{c_1\kappa,\delta} - c_2$.

*Step 3:* The second estimate in (4.31) is obtained by estimating $\mathcal{A}(u)$ from above. We have

$$\mathcal{A}(u) = \int_\Omega u\ln u + \psi u \,\mathrm{d}x \le \max\{\ln u + \psi\}\int_\Omega u\,\mathrm{d}x \le \ln\|u\|_{L^\infty} + \max\psi \le C(1+\|\partial_x u\|).$$

Inserting this into the first estimate of (4.31), the second follows immediately. □

This allows to control the growth of $\mathcal{A}(u)$. Say that $u$ is a solution to the PDE (3.9) and assume that $\ell$ satisfies (4.23) and (4.24), then there must hold for $t_1 < t_2 \in S$

$$\mathcal{A}(u(t_2)) - \mathcal{A}(u(t_1)) \le C_\delta^\psi \int_{t_1}^{t_2} |\dot\ell(t)|\,\mathrm{d}t \le \frac{C_\delta^\psi}{\delta}(t_2 - t_1)\ . \tag{4.32}$$

Given that we know $\mathcal{A}(u(t_1))$, this serves as an upper bound for $\mathcal{A}(u(t_2))$. It only gives a one sided bound, because we have a one sided bound for $\mathcal{D}$ in Theorem 4.2.2. Note that a lower bound for $\mathcal{A}$ exists as well. We know for the case of interest, that solutions are nonnegative and have mean value 1. Thus the term $v\ln v$ has a uniform minimum. This gives a bound for the first part in $\mathcal{A}$. The second part, namely the integral $\int_\Omega v\psi\,\mathrm{d}x$, can also be bounded because $\psi \in \mathrm{W}^{1,2}(\Omega)$ is continuous and must have a minimum.

A bound on $\mathcal{A}$ helps us to deduce bounds for $u$ in $\mathrm{L}^2(\Omega)$, as stated in the next theorem. For this we must use that $\mathbb{L}$ is a bounded linear form on continuous functions.

**Theorem 4.2.4.** *Assume that $\psi \in \mathrm{H}^1(\Omega)$ and $\ell$ satisfies* (4.23)-(4.24). *Furthermore assume that there exists a constant $c_{\mathbb{L}}$, such that for $\mathbb{L}$ in* (3.9) *there holds*

$$|\mathbb{L}(v)| \le c_{\mathbb{L}}\|v\|_{\mathrm{L}^\infty} \qquad \text{for all } v \in \mathrm{H}^1(\Omega). \tag{4.33}$$

*Let the function $u : [0, T_0[ \to \mathrm{H}^1(\Omega)$ be a nonnegative solution to* (3.9). *If there exists a*

*constant $K$ such that*

$$\mathcal{A}(u(t)) \le K \qquad \text{and} \qquad \|u(0)\|_{\mathrm{L}^2} \le K \ , \tag{4.34}$$

*then there exists a constant $C$, depending on $K$ and $\int_\Omega u(0,x)\,dx$, such that the following a priori estimate holds true*

$$\|u(t)\|_{\mathrm{L}^2} \le C \qquad \text{for all } t \in [0, T_0[ \ . \tag{4.35}$$

*Proof.* We want to derive a bound for the growth of $\|u(t)\|_{\mathrm{L}^2}$. We test (3.7) with the solution $u$ itself and get

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u(t)\|_{\mathrm{L}^2}^2 + \|\partial_x u(t)\|_{\mathrm{L}^2}^2 = \Big(\mathbb{L}(u(t)) + \dot{\ell}\Big) \cdot \int_\Omega u(t,x)\partial_x u\,\mathrm{d}x \\ - \int_\Omega \psi'(x)u(t,x)\partial_x u(t,x)\,\mathrm{d}x \ .\end{aligned} \tag{4.36}$$

We derive the following bounds for all addends on the right hand side,

$$\begin{aligned}\Big(\mathbb{L}(u(t)) + \dot{\ell}(t)\Big)\,\mathrm{d}x \int_\Omega u(t,x)\ \partial_x u(t,x)\,\mathrm{d}x &\le 4\Big(\mathbb{L}(u(t)) + \dot{\ell}(t)\Big)\|u(t)\|_{\mathrm{L}^\infty}^2 \ , \\ &\le 4(c_{\mathbb{L}}\|u(t)\|_{\mathrm{L}^\infty} + \frac{1}{\delta})\|u(t)\|_{\mathrm{L}^\infty}^2 \ , \\ -\int_\Omega \psi'(x)\ u(t,x)\ \partial_x u(t,x)\,\mathrm{d}x &\le \frac{1}{4}\int_\Omega (\partial_x u)^2(t,x)\,\mathrm{d}x + \int_\Omega \psi'^2(x)u^2(t,x)\,\mathrm{d}x \ , \\ &\le \frac{1}{4}\|\partial_x u(t)\|_{\mathrm{L}^2}^2 + \|\psi'\|_{\mathrm{L}^2}^2\|u(t)\|_{\mathrm{L}^\infty}^2 \ .\end{aligned}$$

Note that we used $\partial_x(u^2) = 2u\partial_x u$. Inserting these estimates in (4.36) results in

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u(t)\|_{\mathrm{L}^2}^2 + \frac{3}{4}\|\partial_x u(t)\|_{\mathrm{L}^2}^2 \le C_1(1 + \|u(t)\|_{\mathrm{L}^\infty}^3) \ , \tag{4.37}$$

with $C_1$ depending on $\delta$, $c_{\mathbb{L}}$ and $\psi$.

We want to develop from (4.37) a differential inequality with a right hand side depending on $\|u(t)\|_{\mathrm{L}^2}$. Therefor we use the assumed boundedness of $\mathcal{A}$ and the fact that for any $s \in \mathbb{R}_+$, $|s\ln s| \le \frac{2}{\mathrm{e}} + s\ln s$ to derive from (4.34),

$$\|u(t)\ln(u(t))\|_{\mathrm{L}^1} \le \mathcal{A}(u(t)) + \frac{2}{\mathrm{e}} + \|\psi\|_{\mathrm{L}^\infty}\|u(t)\|_{\mathrm{L}^1} \le C_2 \ . \tag{4.38}$$

Since the solution $u$ is nonnegative and has constant mean value, the constant $C_2$ depends on the mean value of $u(0)$ as well as on $K$ and $\psi$.

The step from an a priori bound in an $\mathrm{L}\log\mathrm{L}$-norm to a bound in $\mathrm{L}^2$ is done by employing the following modified version of a Gagliardo-Nirenberg inequality, see Lemma A.2,

$$\forall \varepsilon > 0\ \exists C_\varepsilon\ \forall v \in \mathrm{H}^1(\Omega) : \|v\|_{\mathrm{L}^\infty}^3 \le \varepsilon\|\partial_x v\|_{\mathrm{L}^2}^2 \cdot \|v\ln|v|\|_{\mathrm{L}^1} + C_\varepsilon + C\|v\|_{\mathrm{L}^1}^3 \ . \tag{4.39}$$

Once more we exploit that $\|u(t)\|_{\mathrm{L}^1}$ is constant for all times and combine (4.39), (4.38) and (4.37) to deduce the existence of a constant $C_3$ such that

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u(t)\|_{\mathrm{L}^2}^2 + \frac{3}{4}\|\partial_x u(t)\|_{\mathrm{L}^2}^2 \leq C_3 + \frac{1}{4}\|\partial_x u(t)\|_{\mathrm{L}^2}^2 \ . \tag{4.40}$$

The known mean value $M$, of $u(t)$, allows us to state the existence of constants $C_4$ depending on $M$, such that $\|u(t)\|_{\mathrm{L}^2} - 1 \leq C_4\|\partial_x u\|_{\mathrm{L}^2}$. Thus we get from (4.40) the differential inequality

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u(t)\|_{\mathrm{L}^2}^2 \leq C_3 - \frac{1}{2}\|\partial_x u(t)\|_{\mathrm{L}^2}^2 \leq C_3 + \frac{1}{2C_4} - \frac{1}{2C_4}\|u(t)\|_{\mathrm{L}^2}^2 \ ,$$

which allows us to conclude the desired estimate

$$\|u(t)\|_{\mathrm{L}^2}^2 \leq \max\left\{\|u(0)\|_{\mathrm{L}^2}^2, 2C_4C_3 + 1\right\} \ .$$

□

The a priori estimate in Theorem 4.2.4 and the estimate for the dissipation in Theorem 4.2.2 lead to the existence of solutions for all possible times, provided that (4.23)-(4.24) are satisfied.

**Theorem 4.2.5.** *Let $S = [0, T_0]$ with either finite $T_0 > 0$ or $T_0 = \infty$. Assume $u_0 \in \mathrm{L}^q(\Omega)$ for a $q > 1$ and suppose that $\psi \in \mathrm{H}^1(\Omega)$, $\ell \in \mathrm{W}^{1,\infty}_{loc}(S)$ and $\dot{\ell}$ is piecewise Hölder continuous with Hölder exponent $\alpha > 1/2$. If $\ell$ satisfies on every compact subset $\hat{S} \subset S$ the condition* (4.23)-(4.24) *for a $\delta_{\hat{S}}$, then there exists a solution $u$ to the Problem* (3.9) *on the whole time interval $S$.*

*Proof.* Take any compact interval $\hat{S} = [0, \hat{T}_0] \subset S$. We will show that the solution exists on the whole interval $\hat{S}$. By the arbitrariness of $\hat{T}_0$ it must then exist on the whole interval $S$.

Due to Theorem 4.1.2 we only have to show that the solution stays bounded in $\mathrm{L}^2(\Omega)$. By $u_0 \in \mathrm{L}^q(\Omega)$ we know that the solution is for small positive times $t_0 \in \hat{S}$ in $\mathrm{W}^{1,q}(\Omega)$, thus $u(t_0) \in \mathrm{L}^2(\Omega)$. Fix a time $t_0 \in \hat{S}.$, such $u(t_0) \in \mathrm{L}^2(\Omega)$. Starting from $t_0$ we can guarantee by Theorem 4.1.2 the existence of solutions to the PDE at hand in $\mathrm{W}^{1,2}(\Omega)$ up to a maximal time in $\hat{S}$. We call this time $T_1$.

The inequality (4.32) gives a bound for $\mathcal{A}$ depending on $\hat{T}_0$. Applying Theorem 4.2.4 and the bound for $\mathcal{A}$, we get a bound for $\|u(t)\|_{\mathrm{L}^2}$ for all $t \in [t_0, T_1[$, depending on $\delta_{\hat{S}}$. Thus employing the in Theorem 4.1.2, we have $T_1 = \hat{T}_0$, the solution exists up to the time $\hat{T}_0$. By the arbitrariness of $t_0$, the solution exists in $\left]0, \hat{T}_0\right]$. Finally, $\hat{T}_0 \in S$ was arbitrary and so the solution must exist on the whole interval $S$. □

If the bounds for $\ell$ in (4.23)-(4.24) hold uniformly on $S$ then we get even a uniform $\mathrm{W}^{\beta,2}(\Omega)$ bound for the solution $u$.

**Theorem 4.2.6.** *Let the assumptions of Theorem 4.2.5 hold. Furthermore let $\ell$ satisfy* (4.23)-(4.24) *for a $\delta$ uniformly on the time interval $S$, then the solution $u$ to the PDE* (3.9) *is a priori bounded in $\mathrm{L}^2(\Omega)$ for all times in $S$. Furthermore, for any $\beta \in \left]\frac{1}{2}, 1\right[$ and $\epsilon > 0$, $u(t)$ is bounded in $\mathrm{W}^{\beta,2}(\Omega)$ for all $t \in [\varepsilon, T_0]$.*

*Proof.* From Theorem 4.2.5 we get immediately the existence of a solution on the whole time interval $S$. According to Proposition 4.2.3, there exist positive constants $C$ and $\kappa$ such that the differential inequality

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{A}(u(t)) \leq \mathcal{D}(u(t), \dot{\ell}(t)) \leq -\kappa\mathcal{A}(u(t)) + C, \tag{4.41}$$

holds for all times in $S$. This results in $\mathcal{A}(u(t)) \leq \max\{\kappa A(u(0)), C\}$ for all $t \in S$. The uniform bound for $\mathcal{A}$ allows us to apply Theorem 4.2.4 to derive a uniform bound for $\|u(t)\|_{\mathrm{L}^2}$ for all $t \in S$.

Starting now from any $t \in S$, we can use the achieved uniform bound of $\|u(t)\|_{\mathrm{L}^2}$ to deduce by the estimate (4.15) in Theorem 4.1.2, that for some $\gamma > 0$ and small $\varepsilon > 0$ we have the bound

$$\|u(t+\varepsilon)\|_{\mathrm{W}^{\beta,2}} \leq \epsilon^{-\gamma}(2 + \|u(t)\|_{\mathrm{L}^2}) \ .$$

This bound is then also uniform in time. □

Summarising the results of Lemma 4.2.1, Theorem 4.2.5 and Theorem 4.2.6, we found out that (besides technical conditions like boundedness and piecewise Hölder continuity of $\dot{\ell}$) the condition that the prescribed loading state must not reach one (battery empty) or zero (battery full), is a necessary and sufficient condition for solutions to exist. Thus, in a regime of physical meaningful data we can guarantee the existence of solutions to our model.

## 4.3 Convergence to Steady States

In this section we will show that solutions to the nonlinear Fokker-Planck equation (3.9) converge towards a steady state, provided that the control function $\ell(t)$, see (3.7), converges to a constant value $\ell_* \in ]0, 1[$ as the time $t$ goes to infinity. We know from Lemma 3.3.2, that there exists a whole family of stationary solutions. The value of $\ell_*$ identifies the limit in this family. Of course, we need to guarantee the existence of solutions for all times. For this reason we assume throughout this section on the time

interval $S = [0, \infty[$ that

$$u_0 \in \mathrm{L}^2(\Omega) \qquad u_0 \geq 0 \ , \int_\Omega u_0(x)\,\mathrm{d}x = 1 \ , \tag{4.42}$$

$$\ell \in \mathrm{W}^{1,\infty}(S) \ , \qquad \ell(0) = \int_\Omega x u_0(x)\,\mathrm{d}x \ , \dot{\ell} \text{ is piecewise Hölder continuous} \ , \tag{4.43}$$

$$\exists\ \delta > 0 \ : \ \forall\ t \in S \ : \quad \ell(t) \in [\delta, 1-\delta] \ , \ \|\dot{\ell}\|_{\mathrm{L}^\infty} \leq \frac{1}{\delta} \ , \tag{4.44}$$

$$\psi \in \mathrm{H}^1(\Omega) \tag{4.45}$$

which guarantees the global existence of solutions, according to Theorem 4.2.6.

We will see that, if the control $\ell$ converges to some $\ell_* \in ]0,1[$ in the sense

$$\dot{\ell} \in L^1(]0,\infty[) \ \cap \ L^\infty(]0,\infty[) \quad \text{and hence} \quad \lim_{t\to\infty} \ell(t) = \ell_* \ , \tag{4.46}$$

then the energy $\mathcal{A}$, see (4.25), must converge to some value $A_*$. We see that $A_*$ is the minimum of the energy $\mathcal{A}$ on a convex set $\mathcal{M}_{\ell_*}$. By the strong convexity of $\mathcal{A}$ there is only one such minimiser and it is a stable state. The strong convexity allows us finally to deduce the strong convergence of the minimising sequence to its minimiser. The findings in this section rely heavily on estimates for the dissipation in Theorem 4.2.2 and Proposition 4.2.3.

The first result in this section characterises the minima of $\mathcal{A}$ with the constraint $\mathcal{C}(u) = \ell_*$ as the stationary states described in Section 3.3.

**Proposition 4.3.1.** *For every real number $\ell \in ]0,1[$ there exists a constant $c_\ell$, such that the functional $v \mapsto \mathcal{B}_\ell(v)$ defined as*

$$\mathcal{B}_\ell(v) := \mathcal{A}(v) + c_\ell(\ell - \mathcal{C}(v))$$

*attains its minimum on the set*

$$\mathcal{M} := \{\, u \in L^1(\Omega) \ : \ u \geq 0, \ \int_\Omega u(x)\,dx = 1 \,\}$$

*on exactly one point, namely $U_\ell$ defined in* (3.12)*. Additionally the original functional $v \mapsto \mathcal{A}(v)$ has only one minimiser on the set $\mathcal{M}_\ell := \mathcal{M} \cap \{\, v \in L^1(\Omega) \ : \ \mathcal{C}(v) = \ell\}$ and it is the same $U_\ell$.*

*Proof.* Note that $\mathcal{M}$ and $\mathcal{M}_\ell$ are strongly closed and convex subsets of $L^1(\Omega)$. Moreover, the functional $\mathcal{A}$ is strictly convex as well as its linear disturbed version $\mathcal{B}_\ell$. Hence, there is at most one minimiser of $\mathcal{A}$ on $\mathcal{M}_\ell$ and of $\mathcal{B}_\ell$ on $\mathcal{M}$.

We directly show that $U_\ell$ is in both cases the desired minimiser. The convexity of $u \mapsto u \ln u$ gives

$$\widetilde{u} \ln \widetilde{u} \geq u \ln u + (\ln u + 1)(\widetilde{u} - u) \quad \text{for } u > 0 \text{ and } \widetilde{u} \geq 0.$$

Thus, for all $\widetilde{u} \in \mathcal{M}$ we obtain

$$\begin{aligned}\mathcal{A}(\widetilde{u}) = \int_\Omega \widetilde{u}\ln\widetilde{u} + \psi\widetilde{u}\,\mathrm{d}x \;\geq\; \int_\Omega U_\ell \ln U_\ell + (\ln U_\ell{+}1)(\widetilde{u}{-}U_\ell) + \psi\widetilde{u}\,\mathrm{d}x\ ,\\ \overset{(i)}{=} \mathcal{A}(U_\ell) + \int_\Omega (\beta(\ell)x{+}\ln c_{\beta(\ell)})(\widetilde{u}{-}U_\ell)\,\mathrm{d}x \;\overset{(ii)}{=}\; \mathcal{A}(U_\ell) + \beta(\ell)(\mathcal{C}(\widetilde{u}) - \mathcal{C}(U_\ell)),\end{aligned}$$

where in $(i)$ we used that by the definition of $U_\ell$ in Lemma 3.3.2 we have $\ln U_\ell(x) = \psi(x) + \beta(\ell)x + c_{\beta(\ell)}$ for two constants $\beta(\ell), c_{\beta(\ell)} \in \mathbb{R}$. In $(ii)$ we used $\widetilde{u}, U_\ell \in \mathcal{M}$ and the definition of $\mathcal{C}$, see (3.2). This immediately gives the assertion for $\mathcal{B}_\ell$, when choosing $c_\ell = -\beta(\ell)$. Further applying the definition of $\mathcal{M}_\ell$, results in the assertion for $\mathcal{A}$. □

Next we show that assuming (4.43)-(4.46), the energy $\mathcal{A}$ converges to some value $A_*$. What is important, is that the limit value $A_*$ equals to the minimum of $\mathcal{B}_{\ell_*}$ on $\mathcal{M}$.

**Lemma 4.3.2.** *Suppose that* (4.42)-(4.46) *are satisfied. The solution $u$ to the PDE* (3.9) *then exists on the whole time interval $S = [0, \infty[$ and*

$$\lim_{t\to\infty} \mathcal{A}(u(t)) = A_*\ , \qquad \textit{with} \qquad A_* = \mathcal{A}(U_{\ell_*})\ , \tag{4.47}$$

*where $U_{\ell_*}$ is the stationary state as defined in Lemma 3.3.2.*

*Proof.* We get global existence of the solution $u \in C_{\mathrm{loc}}(S, \mathrm{H}^1_m(\Omega))$ directly from Theorem 4.2.6 together with uniform boundedness of $u(t)$ in $\mathrm{L}^2(\Omega)$ for all $t \in S$. Recall the energy-dissipation (4.27) giving

$$\mathcal{A}(u(t_1)) + \int_0^{t_1} \mathcal{D}(u(t), \dot{\ell}(t))\,\mathrm{d}t = \mathcal{A}(u_0) \quad \text{ for all } t_1 \in S\ . \tag{4.48}$$

The dissipation estimate (4.28) gives

$$\mathcal{D}(u(t), \dot{\ell}(t)) \geq -C|\dot{\ell}(t)| \tag{4.49}$$

for a fixed constant $C$. Thus, the function $\tau \mapsto a(\tau) := \mathcal{A}(u(\tau)) - C\int_0^\tau |\dot{\ell}(t)|\,\mathrm{d}t$ is nonincreasing. By the assumption $\dot{\ell} \in L^1(]0,\infty[)$ and the lower bound $\mathcal{A}(u) \geq (-1/\mathrm{e}{+}\min\psi)$, we know that $a$ is bounded as well. Hence, $a(t) \to a_*$ for $t \to \infty$. The convergence of $a$ results into the convergence $\mathcal{A}(u(t)) \to a_* + C\int_0^\infty |\dot{\ell}(t)|\,\mathrm{d}t =: A_*$.

To deduce the second part of (4.47), we take a look at the dissipation $\mathcal{D}$ along the solution trajectory. The convergence of $\mathcal{A}(u(t))$ in view of (4.48), gives rise to the existence of the integral $\int_0^\infty \mathcal{D}(u(t), \dot{\ell}(t))\,\mathrm{d}t$. Since by (4.49) the negative part of $\mathcal{D}(u(t), \dot{\ell}(t))$ is in $\mathrm{L}^1(S)$, we get that $\mathcal{D}(u(t), \dot{\ell}(t)) \in \mathrm{L}^1(S)$ as well. This means that there must exist $\{t_k\} \subset \mathbb{R}^+$ with $t_k \to \infty$, such that $\mathcal{D}(u(t_k), \dot{\ell}(t_k)) \to 0$ and $\dot{\ell}(t_k) \to 0$. The Proposition 4.2.3 gives that $u$ is uniformly bounded in $\mathrm{H}^1(\Omega)$ and hence, by the continuity of $\mathbb{L}$ in $\mathrm{H}^1(\Omega)$, we deduce

$$\mathcal{D}(u(t_k), 0) = \mathcal{D}(u(t_k), \dot{\ell}(t_k)) + \dot{\ell}(t_k)\mathbb{L}(u(t_k)) \to 0\ , \tag{4.50}$$

as $k \to \infty$. Even more all $\mathcal{D}(u(t_k), 0)$ are finite.

We will now switch over to $v_k := \sqrt{u(t_k)}$. We want to show that $v_k$ is uniformly bounded in $\mathrm{H}^1(\Omega)$. We recall the definition of $\mathcal{D}$ to get

$$\mathcal{D}(u(t_k), 0) = \int_\Omega \frac{\partial_x u(t_k, x))^2}{u(t_k, x)} + 2\psi'(x)\partial_x u(t_k, x) + u(t_k, x)(\psi'(x))^2 \,\mathrm{d}x \;+\; \mathbb{L}(u(t_k))^2 \;. \tag{4.51}$$

By the uniformly boundedness of $u$ in $\mathrm{H}^1(\Omega)$ we know $\mathbb{L}(u(t_k))^2$ is bounded and furthermore $\int_\Omega 2\psi'(x)\partial_x u(t_k, x) + u(t_k, x)(\psi'(x))^2 \,\mathrm{d}x$ exists and must be uniformly bounded as well. Hence, (4.50) and (4.51) give rise to $(\partial_x u(t_k))^2/u(t_k) \in \mathrm{L}^1(\Omega)$. This allows to consider $v_k := \sqrt{u(t_k)}$ as functions in $\mathrm{H}^1(\Omega)$, because $(\partial_x v_k)^2 = (\partial_x u(t_k))^2/u(t_k)$ and furthermore $v_k$ is uniformly bounded in $\mathrm{H}^1(\Omega)$.

We write the dissipation in terms of $v_k$, as in the proof of Theorem 4.2.2, and define $D(v_k) := \mathcal{D}(u(t_k), 0)$, such that

$$D(v_k) := \int_\Omega \Big(2\partial_x v_k + \psi'(x)v_k(x)\Big)^2 \,\mathrm{d}x \;-\; \big[\int_\Omega \Big(2\partial_x v_k + \psi'(x)v_k(x)\Big)\; v_k \,\mathrm{d}x\big]^2 \;.$$

Obviously, $D$ is convex and weakly lower semicontinuous on $\mathrm{H}^1(\Omega)$. Remember, that we work on the weakly closed subset $\mathcal{V} := \{v \in \mathrm{H}^1(\Omega) : \|v\|_{\mathrm{L}^2} = 1\}$ and hence by the Cauchy-Schwarz inequality there holds

$$\int_\Omega \Big(2\partial_x v_k + \psi'(x)v_k(x)\Big)\; v_k \,\mathrm{d}x \le \|2\partial_x v_k + \psi'(x)v_k(x)\|_{\mathrm{L}^2} \;, \tag{4.52}$$

which results in $D(v) \ge 0$ for all $v \in \mathcal{V}$. Furthermore we know on $\mathcal{V}$ that $D(v) = 0$, if and only if for some factor $c \in \mathbb{R}$ there holds $c\, v = 2\partial_x v + \psi'(x)v(x)$ almost everywhere. Consequently, this implies that $v = a\mathrm{e}^{(bx-\psi(x))/2}$ for some constants $a, b \in \mathbb{R}$.

We summarise that $v_k$ is uniformly bounded in $\mathrm{H}^1(\Omega)$ and thus converges weakly to some $v_* \in \mathrm{H}^1(\Omega)$ when passing to a subsequence. We also know that $\|v_k\|_{\mathrm{L}^2} = \|v_*\|_{\mathrm{L}^2} = 1$. Additionally, $D$ is weakly lower semicontinuous and convex and $D(v_k) \to 0$, which is the infimum of $D$ on the convex set $\{v \in \mathrm{H}^1(\Omega) :\; \|v\|_{\mathrm{L}^2} = 1\}$. Thus, by the lower semicontinouity of $D$, there must hold $D(v_*) = 0$. And so $v_* = a\mathrm{e}^{(bx-\psi(x))/2}$ for some constants $a, b \in \mathbb{R}$. To identify $a$ and $b$ we exploit $\ell_* = \lim_k \mathcal{C}(v_k^2)$. By the weak convergence of $v_k$ in $\mathrm{H}^1(\Omega)$ we have strong convergence of $v_k^2 = u(t_k) \to v_*^2 = u_*$ in $\mathrm{C}(\bar{\Omega})$ and so we get $\mathcal{C}(v_*^2) = \ell_*$. Together with $\|v_*\|_{\mathrm{L}^2} = \|u_*\|_{\mathrm{L}^1} = 1$ and the positivity of $u_*$, all these properties of $u_* = v_*^2$ fix $a$ and $b$, see Lemma 3.3.2, and so finally $u_* = U_{\ell_*}$.

Now that we have established $u(t_k) \to U_{\ell_*}$ in $\mathrm{C}(\bar{\Omega})$, we deduce for the limit energy value $A_* = \mathcal{A}(U_{\ell_*})$, which results from the continuity of $\mathcal{A}$. □

In the proof of the last lemma we showed that along a special sequence of times, the solution converges to the uniquely defined stationary solution $U_{\ell_*}$. Further using that we have identified the limit energy $A_*$, we can derive the convergence of the whole evolution $u(t)$ for $t \to \infty$. This is stated in our final convergence theorem.

**Theorem 4.3.3.** *We again assume that* (4.42)-(4.46) *are satisfied. The function* $u : S \to \mathrm{H}^1(\Omega)$*, which is a solution to the PDE* (3.9) *then converges in the sense*

$$u(t) \to U_{\ell_*} \qquad \text{in } \mathrm{L}^2(\Omega) \ ,$$

*where* $U_{\ell_*}$ *is the stationary state as described in Lemma 3.3.2.*

*Proof.* Take an arbitrary sequence $\{t_k\}_k \subset S$ with $t_k \to \infty$ as $k \to \infty$. By Theorem 4.2.6 we have a uniform bound on $u(t_k)$ in all $\mathrm{W}^{\theta_1,2}(\Omega)$ for any $\theta_1 \in [0,1[$. Using the compact embedding of $\mathrm{W}^{\theta_1,2}(\Omega) \hookrightarrow \mathrm{W}^{\theta_2,2}(\Omega)$, for all $\theta_2 < \theta_1$, we get a $u_* \in \mathrm{W}^{\theta_2,2}(\Omega)$, such that along a subsequence there must hold $u(t_k) \to u_*$ in $\mathrm{W}^{\theta_2,2}(\Omega)$. For ease of notation we do not relabel the sequence $t_k$.

In the next step we show that $u_* = U_{\ell_*}$. We consider the functional $\mathcal{B}_{\ell_*}$ and the set $\mathcal{M}$ as defined in Proposition 4.3.1. We know that $\mathcal{C}(u(t_k)) = \ell(t_k) \to \ell_*$ and by Lemma 4.3.2 even $\mathcal{A}(u(t_k)) \to A(U_{\ell_*})$, with $U_{\ell_*}$ the stable solution from Lemma 3.3.2. Since $\mathcal{B}_{\ell_*}(u) = \mathcal{A}(u) + c\big(\ell_* - \mathcal{C}(u)\big)$ we thus have

$$\mathcal{B}_{\ell_*}(u(t_k)) \to \mathcal{A}(U_{\ell_*}) = \mathcal{B}_{\ell_*}(U_{\ell_*}) = \min_{v \in \mathcal{M}} \mathcal{B}(v) \ , \tag{4.53}$$

and so $u(t_k) \in \mathcal{M}$ is a minimising sequence. We have that $\mathcal{M} \in \mathrm{L}^1(\Omega)$ is a closed convex set and $\mathcal{B}_{\ell_*}$ has the form $\mathcal{B}_{\ell_*}(v) = \int_\Omega \Psi(v(x)) + f(x)v(x)\,\mathrm{d}x + c$ with $\Psi(\xi) = \xi \log \xi$. So we can use a result by Visintin, [40, Thrm. 8] stating that $u(t_k) \to U_{\ell_*}$ in $\mathrm{L}^1(\Omega)$, because $U_{\ell_*}$ is the minimiser of $\mathcal{B}_{\ell_*}$ on $\mathcal{M}$. The limit of the already achieved stronger convergence must then be the same. As the possible limit of all bounded sequences is unique, we have convergence of the whole family $u(t)$ in all spaces $\mathrm{W}^{\theta_2}(\Omega)$ with $\theta_2 \in [0,1[$. □

# 5 Numerical Analysis of the Nonlinear Fokker–Planck Equation

This chapter shows the behaviour of solutions to the PDE (3.9) using numerical simulations. In the first section we will describe the numerical method which we employed. It is a standard finite element method, using piecewise linear ansatz functions in one spatial dimension. In order to achieve higher accuracy, the nonlinearity of the PDE is resolved implicitly. We will not describe the method of finite elements in general, nor will we discuss or show details of actual source code of the implementation. For a detailed description of the method of finite elements as a numerical approximation technique for PDEs, we refer to the textbook by Ciarlet [7]. The second section is devoted to the convergence of numerical solutions to the exact solutions. We formulate the convergence result in a more general way, so that it might be applied to other settings of discretised nonlinear PDEs as well. In the third section we study the qualitative behaviour of solutions in numerical experiments. We will discuss the hysteretic behaviour and the way the parameters $\tau$ and $\nu$ influence it.

## 5.1 Numerical Method

In order to cast the Fokker–Planck equation in the framework of conforming finite elements, we write the weak formulation of (3.9). We use the notation $(\cdot,\cdot)_{\mathrm{L}^2}$ for the usual scalar product in the Hilbert space $\mathrm{L}^2(\Omega)$. The weak formulation of the PDE on a time interval $S := ]0,T]$ then reads

$$\forall t \in S\ ,\ \forall v \in H^1(\Omega) \tag{5.1}$$
$$\Big(\partial_t u(t)\ ,\ v\Big)_{\mathrm{L}^2} = -\frac{\nu^2}{\tau}\Big(\partial_x u(t)\ ,\ \partial_x v(t)\Big)_{\mathrm{L}^2} - \frac{1}{\tau}\Big([\psi' - \tau p(t) - \mathbb{L}(u(t))]u(t)\ ,\ \partial_x v(t)\Big)_{\mathrm{L}^2}\ ,$$

together with a given initial value $u_0 \in \mathrm{L}^2(\Omega)$, $u_0 \geq 0$, $\int_\Omega u_0(x)\,\mathrm{d}x = 1$ such that $u(0) = u_0$. As in Chapter 3, $\mathbb{L}$ is a linear functional as defined in (3.8) and the given data are $\psi \in \mathrm{H}^1(\Omega)$, $p(t) \in \mathrm{C}^\alpha(\Omega)$ and $\alpha > 1/2$. Furthermore the function $\ell(t) := \int_\Omega x\, u_0(x)\,\mathrm{d}x + \int_0^t p(s)\,\mathrm{d}s$ shall have at all times values in the unit interval, i.e. $\ell(t) \in ]0,1[$, which guarantees the existence of solutions on $S$, according to Section 4.2.

We carry the weak formulation over to a finite dimensional subspace $V_h \subset \mathrm{H}^1(\Omega)$ with $V_h = \mathrm{span}\{\varphi_1, \cdots, \varphi_n\}$. The subscript $h$ emphasises that this space shall rise from a space discretisation and is connected with a spatial grid size $h$. In our implementation we used $V_h$ as the space of piecewise linear functions on an equidistant grid on the interval $[0,1]$. However, the following reasoning works with other choices for finite dimensional

subspaces $V_h$ and does not depend on this choice. The restriction of the weak formulation to the subspace $V_h$ results in a system of the form

$$M\dot{U}(t) = -\frac{\nu^2}{\tau}AU(t) - \frac{1}{\tau}\widehat{B}U(t) + p(t)BU(t) + \Big(L\ ,\ U(t)\Big)_{\mathrm{e}}BU(t)\ , \tag{5.2}$$

where we write $(\cdot,\cdot)_{\mathrm{e}}$ for the Euclidean scalar product of two vectors in $\mathbb{R}^N$. The vector $L \in \mathbb{R}^N$ and the matrices $M, A, B, \widehat{B} \in \mathbb{R}^{N\times N}$ are defined as

$$M_{i,j} := \int_\Omega \varphi_i(x)\varphi_j(x)\,\mathrm{d}x\ , \qquad A_{i,j} := \int_\Omega \nabla\varphi_i(x)\nabla\varphi_j(x)\,\mathrm{d}x\ , \tag{5.3}$$

$$B_{i,j} := \int_\Omega \varphi_j(x)\nabla\varphi_i(x)\,\mathrm{d}x\ , \qquad \widehat{B}_{i,j} := \int_\Omega \psi'(x)\varphi_j(x)\nabla\varphi_i(x)\,\mathrm{d}x\ , \tag{5.4}$$

$$L_i := \frac{1}{\tau}\mathbb{L}(\varphi_i)\ . \tag{5.5}$$

The matrix $M$ is invertible and so the ODE has, supplying an initial value, a unique solution up to a possible blow up time, as the right hand side is quadratic and thus locally Lipschitz continuous. The ODE in (5.2) is then solved, using the trapezoid time stepping method. In the context of finite element methods, this is called a Cranck-Nicholson scheme. We define some notation

| | |
|---|---|
| $U^j \in \mathbb{R}^N\ ,$ | approximation of the solution at a time $t_j\ ,$ |
| $k_j := t_j - t_{j-1}\ ,$ | time step size , |
| $S^j := -\frac{\nu^2}{\tau}A - \frac{1}{\tau}\widehat{B} + p(t_j)B\ .$ | |

The Crank-Nicholson time-stepping scheme means at every time step we solve the following equation for $U^j$ given $U^{j-1}$,

$$\frac{1}{k_j}M(U^j - U^{j-1}) = \frac{1}{2}(S^jU^j + S^{j-1}U^{j-1} + \Big(L\ ,\ U^j\Big)_{\mathrm{e}}BU^j + \Big(L\ ,\ U^{j-1}\Big)_{\mathrm{e}}BU^{j-1})\ ,$$

or the equivalent form with $f_j := MU^{j-1} + \frac{1}{2}k_j[S^{j-1}U^{j-1} + \Big(L\ ,\ U^{j-1}\Big)_{\mathrm{e}}BU^{j-1}]$

$$g(U^j) := (M - k_j\frac{1}{2}S^j)U^j - k_j\frac{1}{2}(L\ ,\ U^j)_{\mathrm{e}}BU^j - f_j = 0\ . \tag{5.6}$$

Choosing $k_j$ close to zero guarantees that a solution to this nonlinear equation exists. This can be seen with the help of the implicit function theorem and the invertibility of $M$. We solve equation (5.6) by employing a usual Newton method. The iterative scheme for a sequence $U_i^j \to U^j$ reads

$$U_i^j = U_{i-1}^j - \mathrm{D}g(U_{i-1}^j)^{-1}g(U_{i-1}^j)\ , \tag{5.7}$$

with

$$\mathrm{D}g(U_{i-1}^j) = M - k_j \frac{1}{2} S^j - \frac{1}{2} k_j (L\ ,\ U_{i-1}^j)_{\mathrm{e}} B - \frac{1}{2} k_j B U_{i-1}^j \otimes L\ .$$

By $a \otimes b$ we denote the matrix $(a_i b_j)_{i,j}$. The Matrix $\mathrm{D}g(U_i^j)$ is invertible for $k_j$ small enough. This is true since it is a disturbance of the invertible matrix $M$.

The Matrices $M$, $S^j$ and $B$ are sparse, but in general $BU_{i-1}^j \otimes L$ is dense. In practice it even turns out, that this term is a fully occupied matrix and so is $\mathrm{D}g(U_i^j)$. This is eventually a bottle neck in terms of speed for numerical calculations. Sparse matrices would allow the usage of well established fast computational algebraic algorithms for solving such equations. In order to avoid inverting a dense matrix, we make explicit use of the form of $\mathrm{D}g(U_{i-1}^j)$. It is a sparse matrix plus a rank one matrix of the form $BU_{i-1}^j \otimes L$. For this we use the simple identity which holds for any invertible $H \in \mathbb{R}^{N\times N}$ and $a, b, x \in \mathbb{R}^N$, namely

$$(H + a \otimes b)^{-1} x = H^{-1} x - \frac{(b, H^{-1}x)}{1 + (b, H^{-1}a)} H^{-1} a\ .$$

Note that this imposes an implicit condition when $(H + a \otimes b)$ is actually invertible, by requiring the denominator on the right hand side, to be different from zero. In the resulting calculation we only need to invert $H$, which is in our case a sparse matrix. In our simulations this results in a tremendous reduction of computation time.

## 5.2 Convergence of Numerical Solutions

In this section we will show that the discrete numerical solutions converge to the exact solution of our actual problem. The existing literature provides numerical convergence results for semilinear parabolic problems, but they do not seem to fit for our problem. One can easily find results for reaction-diffusion equations of the form $\partial_t u = \triangle u + f(u)$ and $f : \mathbb{R} \to \mathbb{R}$, see [21]. The key difference to our setting is that in the case of reaction-diffusion equations $f$ maps from $\mathrm{L}^2$ to $\mathrm{L}^2$ when interpreted as an operator acting on functions. On the other hand the nonlinearity in our consideration, see (3.9), maps from $\mathrm{H}^\theta$, $\theta > 1/2$, to $\mathrm{H}^{-1}$. It is also hard to adapt techniques, because we cannot provide higher regularity of the exact solution in space or time. Building on standard techniques for linear parabolic problems, see the book by Larsson and Thomée [26], the usual approach is to use the regularity of the exact solution for interpolation estimates. We can achieve more regularity in time if we assume more regularity of the given date $\ell$, which is the prescribed first moment of the solution in time. On the other hand, more regularity in space needs more regularity of the energy $\psi$, whose derivative $\psi'$ appears as a spatial varying coefficient in the PDE at hand, (3.9). Our explicit aim is to include functions $\psi'$ which are singular at the boundary and are in $\mathrm{L}^2(\Omega)$ but not in $\mathrm{H}^1(\Omega)$. So the assumption of better regularity on $\psi'$ seems not reasonable, and we do not get better spatial regularity of the exact solution.

Our arguments are of abstract form. We are able to state that for a sequence of space discretisation fineness $h_j \to 0$ and a sequence of time step-size $k_j \to 0$ the numerical solutions converge to the exact solution in the space of continuous functions. For this $h_j$ and $k_j$ must vanish together in a suitable way. As our arguments are abstract and not constructive, we lack to say what this suitable way might be. Nevertheless, the achieved convergence result gives confidence that our numerical solutions are related to the exact solutions of the PDE at hand. This allows us to use numerical experiments in order to make statements on the evolutionary behaviour of the solution depending on the given data.

We deduce the desired convergence in two steps. First, we show convergence of the semi-discrete system, (5.2), which is a discretisation in space only, but not in time. To do so, we derive a priori bounds in spaces $\mathrm{H}^\theta$, $\theta \in [0,1]$ which do not depend on the discretisation parameter and so for $h_j \to 0$ at least a subsequence of solutions $U_{h_j}$ must converge weakly. We identify the limit as the unique exact solution of the original problem. The uniqueness of the limit gives the weak convergence of the whole sequence. In the second step we carry this convergence over to the fully discretised problem by employing standard arguments for the numerical integration of ODEs in $\mathbb{R}^n$. Thus, we carry the convergence of the semi-discretised problem over to the fully discretised one.

We do not derive convergence speeds or dependencies of space and time discretisation. The difficulty is here, the convergence of the semi-discrete problem when discretised only in space. If one is able to improve the convergence results presented here and derive a priori estimates for the convergence of the semi-discrete solutions, then one could use techniques of Lubich and Ostermann [27] to further derive estimates for the error of the time discretisation. Another work should be mentioned, because it motivated the use of semigroup techniques to show numerical convergence, it is a paper by Bakaev [2]. The author exploits, for linear parabolic problems, the connection of the discrete operators to their undiscretised counterparts. Considering only linear problems, Bakaev is able to derive error estimates for the discrete semigroup and even for the fully discretised problem.

The following statements are formulated in a more general setting, such that its treatment can be applied to other suitable problems. We consider a complex Hilbert space X with norm $\|\cdot\|_{\mathrm{X}}$ and scalar product $(\cdot,\cdot)_{\mathrm{X}}$ and employ complex operator theory. Furthermore we denote by $V$ a dense subspace of X, which is also a Hilbert space equipped with norm $\|\cdot\|_1$. By $V^{-1}$ we denote the dual of $V$ with norm $\|\cdot\|_{-1}$. We then have the cascade of canonical embeddings

$$V \hookrightarrow \mathrm{X} = \mathrm{X}^* \hookrightarrow V^{-1} \ .$$

**Assumption 5.2.1.** *Let $-A : D(A) \subset V^{-1} \to V^{-1}$ be a linear and sectorial operator with compact resolvent and dense domain $D(A) = V$ in $V^{-1}$, such that there exist*

*constants $m > 0$, $\omega < 0$ and $\theta \in ]\pi/2, \pi[$, satisfying*

$$\rho(-A) \supset S_{\theta,\omega} := \{\lambda \in \mathbb{C} : \lambda \neq \omega, |\arg(\lambda - \omega)| \leq \theta\} ,$$
$$\|R(\lambda, -A)v\|_{-1} \leq \frac{m}{|\lambda - \omega|}\|v\|_{-1} \qquad \text{for all } v \in V, \lambda \in S_{\theta,\omega} .$$

*By $\rho(-A)$ we denote the resolvent set of $-A$. Furthermore we assume, that $A$ is self-adjoint and gives the norm and scalar product in $V$ by*

$$\|x\|_1 := \sqrt{(x, \overline{x})_1} , \qquad (x, y)_1 := \langle Ax, \overline{y} \rangle ,$$

*which results into the identity $\|x\|_1 = \|Ax\|_{-1}$. The operator $A$ shall also be weakly closed. This means, that for any sequence $v_j \in V$ such that $AV_j \rightharpoonup w$ in $V^{-1}$, there must exist $v \in V$ such that $v_j \to v$ in $V$ and $Av = w$.*

Note that a simple consequence of $m > 0$ and $\omega < 0$ is the existence of a constant $c_A$, such that

$$\|x\|_1 \geq c_A\|x\|_{-1} .$$

Whenever calculations will depend on $m$, $\omega$, $\theta$ or $c_A$, we say that they depend on the sectorial properties of $A$.

With this notion we can also define intermediate spaces for $\theta \in [-1, 1]$ as the domains of fractional powers of $A$, by

$$V^\theta = D(A^{\frac{\theta+1}{2}}) \qquad \text{with norm: } \|x\|_\theta := \|A^{\frac{1+\theta}{2}}x\|_{-1} , \tag{5.8}$$

and we assume that they are compactly embedded into each other, such that

$$V^{\theta_2} \hookrightarrow V^{\theta_1} \qquad \text{for all } 0 \leq \theta_1 \leq \theta_2 \leq 1 . \tag{5.9}$$

This compact embedding can actually be deduced from the assumption of compact resolvents. The space $V^0$ plays a special role, for which we need the following assumption.

**Assumption 5.2.2.** *The intermediate space $V^0$, defined as in* (5.8)*, must satisfy*

$$V^0 = \mathrm{X}$$

*with equivalence of norms.*

The setting we work on, is $\mathrm{X} = \mathrm{L}^2(\Omega)$ and $V^1 = \mathrm{H}^1_m(\Omega)$ and $A = \frac{\nu^2}{\tau}\partial_{xx}$. By $\mathrm{H}^1_m(\Omega)$ we denote the space of mean-value free functions $v \in \mathrm{H}^1(\Omega)$. The method works in other settings as well and we therefor keep our results in a general form.

We also introduce a general form of a discretisation of $A$ on a finite dimensional subspaces. For $n \in \mathbb{N}$ and a set of linear independent elements $\{\varphi_j\}_j \subset V$, we define

$$V_h = \{\varphi_1, \varphi_2, \cdots, \varphi_n\} \subset V , \tag{5.10}$$

which is, as constructed, isomorph to $\mathbb{C}^n$. We define the matrices $A_h$, $M_h$ and $D_h$ as

$$A_{hi,j} = \langle A\varphi_i, \varphi_j \rangle \ , \qquad M_{hi,j} = (\varphi_i, \varphi_j)_{\mathrm{X}} \ , \qquad D_h = M_h^{-1} A_h \ . \tag{5.11}$$

Note that $M_h$, $A_h$ and $D_h$ are invertible, symmetric and positive definite matrices. First, we introduce some notation. We define suitable norms in the finite dimensional subspaces and give relations to norms in spaces, in which they are embedded. Let $v \in \mathbb{C}^n$, then we define the embedding operator

$$T_h : \mathbb{C}^n \to \operatorname{span}\{\varphi_1, \varphi_2, \cdots, \varphi_N\} \subset V \ , \qquad T_h v = \sum_{j=1}^{N} v_j \varphi_j \ . \tag{5.12}$$

In other sections we will write also $v$ instead of $T_h v$ and will not distinguish between both forms in our notation. In this section, however, we distinguish them, since the switch from $v$ to $T_h v$ is essential for the following thoughts.

Next to $T_h$, we introduce the projection $P_h : V^{-1} \to \mathbb{C}^n$, which is for any given $w \in V^{-1}$ defined as the solution of

$$\langle w, T_h v \rangle = (P_h w, M_h v)_{\mathrm{e}} \qquad \text{for all } v \in \mathbb{C}^n \ . \tag{5.13}$$

By the identity $(T_h P_h w, T_h v)_{\mathrm{X}} = (P_h w, M_h v)_{\mathrm{e}}$, we see that $T_h P_h$ is the X-orthogonal projection from $V$ into $V_h$. We need that the projection $P_h$ is stable in the $V^1$ norm. The precise meaning of stability is formulated in the following assumption.

**Assumption 5.2.3.** *There exists a constant $c_P$*

$$\|P_h v\|_1 \le c_P \|v\|_1 \ . \tag{5.14}$$

*The crucial point is, that $c_P$ does in general depend on the choice or construction of the discrete space $V_h$. Consequently, the inequality* (5.14) *imposes extra conditions when considering a family of discrete spaces $V_{h_j}$, namely that there is a $c_P$ independent of $j$. For sufficient conditions to satisfy this stability for the* $\mathrm{L}^2$ *projection, we refer to Bramble, Pasciak and Steinbach [4].*

We denote the usual Euclidean norm by $\|\cdot\|_{\mathrm{e}}$ and define for $v \in \mathbb{C}^n$ discrete versions of norms,

$$\|v\|_{h,-1} := \|M_h A_h^{-\frac{1}{2}} v\|_{\mathrm{e}} \qquad = \sup_{w \in V_h, \|w\|_1} (T_h v, w)_{\mathrm{X}} \qquad \le \|T_h v\|_{-1} \ , \tag{5.15}$$

$$\|v\|_{h,0} := \|D_h^{\frac{1}{2}} v\|_{h,-1} \qquad = \|M_h^{\frac{1}{2}} v\|_{\mathrm{e}} \qquad = \|T_h v\|_{\mathrm{X}} \ , \tag{5.16}$$

$$\|v\|_{h,1} := \|D_h v\|_{h,-1} \qquad = \|A_h^{\frac{1}{2}} v\|_{\mathrm{e}} \qquad = \|T_h v\|_1 \ . \tag{5.17}$$

The discrete norms $\|\cdot\|_{h,0}$ and $\|\cdot\|_{h,1}$ are standard for this type of methods. Those two norms result simply from the mapping $T_h$. The more interesting definition is the norm $\|v\|_{h,-1}$. It stems from the question of how a vector $v \in \mathbb{C}^n$ can be interpreted as

an element of the dual space of $V_h^1 := (\mathbb{C}^n, \|\cdot\|_{h,1})$. Such an interpretation is defined by the action of $w \in \mathbb{C}^n$ on another element $v \in V_h^1$. At first glance one might choose an interpretation of the form $\langle w, v\rangle = (w, v)_{\mathrm{e}}$, but this would not fit our needs. Only the choice $\langle w, v\rangle := (w, M_h v)_{\mathrm{e}} = (T_h w, T_h v)_{\mathrm{X}}$ leads to the identities in (5.16) and (5.17). This allows us to express the norms $\|\cdot\|_{h,0}$ and $\|\cdot\|_{h,1}$ by $\|\cdot\|_{h,-1}$ in powers of $D_h$. This choice also mimics the embedding $V \hookrightarrow X \hookrightarrow V^{-1}$. Furthermore these definitions allow us to fit the discretised setting in the context of sectorial operators. Imitating the notation for the original space $V$ and $V^{-1}$, we will use the following names for $\mathbb{C}^n$ when equipped with the above norms

$$V_h^{-1} := (\mathbb{C}^n, \|\cdot\|_{h,-1}) \ , \qquad V_h^0 := (\mathbb{C}^n, \|\cdot\|_{h,0}) \ , \qquad V_h^1 := (\mathbb{C}^n, \|\cdot\|_{h,1}) \ .$$

Note that $\|v\|_{h,-1}$ is only bounded from above by $\|T_h v\|_{-1}$. By the assumed stability of the projection $P_h$ in (5.14), we can achieve a bound from below. Take any $w \in V_h^{-1}$, then

$$\begin{aligned} \|T_h w\|_{-1} &= \sup_{v\in V, \|v\|_1} |(T_h w, v)_{\mathrm{X}}| = \sup_{v\in V, \|v\|_1} |(w, M_h P_h v)_{\mathrm{e}}| = \sup_{v\in V, \|v\|_1} |\langle w, P_h v\rangle| \\ &\leq \sup_{v\in V, \|v\|_1} \|w\|_{V_h^{-1}} \, \|P_h v\|_{V_h^1} \leq \|w\|_{V_h^{-1}} c_P \ . \end{aligned} \tag{5.18}$$

Now we state the general form of problems and their discretised versions, for which we want to show the desired convergence of the approximative solutions. We define the finite time interval $S = ]0, T_0]$ and let $N$ be a mapping from $V^\gamma \times \mathbb{R}$ into $V^{-1}$ for a $\gamma \in [-1, 1[$. Then we seek for solutions to the operator differential equation

$$\partial_t u(t) = -Au(t) + N(u(t), t) \ , \qquad u(0) = u_0 \ , \tag{5.19}$$

for some $u_0 \in V^{-1}$. Its discretised version is an ODE in $\mathbb{C}^n$,

$$\partial_t U(t) = -D_h U(t) + P_h N(T_h U(t), t) \ , \qquad U(0) = Q_h u_0 \ , \tag{5.20}$$

with a suitable mapping $Q_h : X \to \mathbb{C}^n$. Thus, the discretised version of $N$ is $P_h N(T_h \cdot, \cdot)$, which maps : $\mathbb{C}^n \times \mathbb{R}$ into $\mathbb{C}^n$. The possibly nonlinear mapping $N$ must satisfy the following assumption.

**Assumption 5.2.4.** *Let $\gamma \in ]0, 1[$. The mapping $N : V^\gamma \times \mathbb{R} \to V^{-1}$ is locally Lipschitz in the first argument and Hölder continuous with Hölder exponent $\alpha_N > \frac{1}{2}$ in the second. This means there exists an increasing function $g$, such that for all $v_1, v_2 \in V^\gamma$ and $s_1 < s_2 \in \mathbb{R}$,*

$$\|N(v_1, s_1) - N(v_2, s_2)\|_{-1} \leq g(\|v_1\|_\gamma + \|v_2\|_\gamma)\Big(\|v_1 - v_2\|_\gamma + |s_1 - s_2|^{\alpha_N}\Big) \ . \tag{5.21}$$

*We also assume that there exist constants $c_1 \geq 0$, $c_2 \geq 0$, such that for all $s \in S$ and*

*$v \in V^\gamma$ there holds*

$$\|N(v,s)\|_{-1} \le c_1 + c_2\|v\|_\gamma \ . \tag{5.22}$$

This assumption guarantees the existence of solutions for both problems, (5.19) and (5.20), since it permits to deduce a priori bounds to the solution for any finite time. In case of the Fokker-Planck equation (3.9) we can not satisfy (5.22). However, we state this as a necessity for the general case and will later see that we can modify the original problem to fit (5.22) without changing the actual problem.

In order to derive a priori bounds of solutions to (5.20), which are independent of the specific choice of $V_h$, we use the notion of sectorial operators. We transform techniques from the existence proof of the exact solution, see Theorem 4.1.2, to the semi-discretised setting. For this we need knowledge about the discretised operator as stated in the following lemma. For our implemented method we could use known results for the eigenvalues of tridiagonal matrices, as they appear in the stiffness matrices for the 1-D Laplacian, see [42]. In order to keep the results applicable for other problems as well, we derive properties of the discretised elliptic operator in a rather abstract form.

**Lemma 5.2.5.** *The matrix $-D_h$, as defined in* (5.11)*, considered as an operator*

$$D_h : V_h^{-1} \to V_h^{-1} \ ,$$

*is sectorial. That means, that there exist constants $\theta \in ]\pi/2, \pi[$, $m > 0$ and $\omega \in \mathbb{R}$ such that*

$$\begin{aligned} &\rho(-D_h) \supset S_{\theta,\omega} := \{\lambda \in \mathbb{C} : \lambda \neq \omega, |\arg(\lambda - \omega)| \le \theta\} \\ &\|R(\lambda, -D_h)v\|_{h,-1} \le \frac{m}{|\lambda - \omega|}\|v\|_{h,-1} \qquad \textit{for all } v \in \mathbb{C}^n, \lambda \in S_{\theta,\omega}. \end{aligned}$$

*The constants $\theta$, $m$ and $\omega$ are independent from the exact choice of $V_h$ and depend on the sectorial properties of $-A$.*

*Proof.* First note that $D_h$ is, due to its construction in (5.11), a symmetric and positive definite matrix. Thus, all its eigenvalues are real and positive. Its minimal eigenvalue $\lambda_{h,0}$ satisfies

$$\begin{aligned} \lambda_{h,0} &= \inf_{v\in\mathbb{C}^n} \frac{\|D_h v\|_{\mathrm{e}}}{\|v\|_{\mathrm{e}}} = \inf_{v\in\mathbb{C}^n} \frac{\|D_h A_h^{\frac{1}{2}} v\|_{\mathrm{e}}}{\|A_h^{\frac{1}{2}} v\|_{\mathrm{e}}} = \inf_{v\in\mathbb{C}^n} \frac{\|D_h v\|_{h,-1}}{\|v\|_{h,-1}} = \inf_{v\in\mathbb{C}^n} \frac{\|T_h v\|_1}{\|v\|_{h,-1}} \ , \\ &\ge c \inf_{v\in\mathbb{C}^n} \frac{\|T_h v\|_1}{\|T_h v\|_{-1}} \ge c \inf_{v\in V^1} \frac{\|v\|_1}{\|v\|_{-1}} =: C_A > 0 \ . \end{aligned}$$

Obviously, $C_A$ does not depend on the choice of the subspace $V_h$, but on the sectorial properties of $A$. Next we exploit the inequality $C_A\|D_h v\|_{h,-1} \le \|v\|_{h,-1}$. Let us denote by $(\cdot,\cdot)_{h,-1}$ the sesquilinear form associated with $\|\cdot\|_{h,-1}$. Then we get for all $\lambda \in \mathbb{C}$

with $\mathrm{Re}\lambda \geq -C_A/2$ and $v \in \mathbb{C}^N$,

$$\begin{aligned}\|(\lambda + D_h)v\|_{h,-1}^2 &= |\lambda|^2\|v\|_{h,-1}^2 + 2\mathrm{Re}(\lambda)\big(D_h v, v\big)_{h,-1} + \|D_h v\|_{h,-1}^2 \ , \\ &\geq \|v\|_{h,-1}^2(\mathrm{Im}(\lambda)^2 + \mathrm{Re}(\lambda)^2 + 2C_A\mathrm{Re}(\lambda) + C_A^2) \ , \\ &\geq \|v\|_{h,-1}^2[(\mathrm{Re}(\lambda) + C_A)^2 + \mathrm{Im}(\lambda)^2] \geq \|v\|_{h,-1}^2[\mathrm{Re}(\lambda)^2 + \mathrm{Im}(\lambda)^2] \ , \\ &\geq \|v\|_{h,-1}^2|\lambda|^2 \ .\end{aligned}$$

Thus, we deduce that $(\lambda + D_h)$ is invertible for all $\lambda \in \mathbb{C}$ with $\mathrm{Re}\lambda \geq -C_A/2$ and furthermore for $R(\lambda, -D_h) = (\lambda + D_h)^{-1}$ we get

$$\|\lambda R(\lambda, -D_h)v\|_{h,-1} \leq \|v\|_{h,-1} \qquad \text{for all } \lambda \in \mathbb{C},\ \mathrm{Re}\lambda \geq 0 \ .$$

According to [28, Prop.2.1.11], this estimate is sufficient for $D_h$ to be sectorial. Tracking the constants in the proof therein, we identify $\omega = -C_A/2 < 0$, $m = 2$ and $\theta = \frac{3}{4}\pi$. □

This insight allows us to use semigroup techniques for further estimates. In order to do this we need to connect the norm in $V^\theta$, $\theta \in [-1, 1]$, for which we want to derive bounds, with the norms $\|\cdot\|_{h,-1}$, $\|\cdot\|_{h,0}$ and $\|\cdot\|_{h,1}$, which belong to the mentioned semigroup in $\mathbb{C}^n$. We do so by using that $\|T_h \cdot\|_1 = \|\cdot\|_{h,1}$ and the estimate for $\|T_h \cdot\|_{-1}$ in (5.18). Thus for $\theta \in [-1, 1]$ there holds

$$\|T_h v\|_\theta \leq c\|T_h v\|_{-1}^{1-\theta}\|T_h v\|_1^\theta \leq c\|v\|_{h,-1}^{1-\theta}\|v\|_{h,1}^\theta \ , \qquad \text{for all } v \in \mathbb{C}^n \ . \tag{5.23}$$

One identity, which we will exploit extensively, is the variation of constants formula. This formula is satisfied for $U$, the solution to (5.20), and for all $t$ in the existence interval of the solution. It reads

$$U(t) = \mathrm{e}^{-D_h t}U(0) + \int_0^t \mathrm{e}^{-D_h(t-s)}P_h N(T_h U(s), s)\,\mathrm{d}s \ . \tag{5.24}$$

Even though $\mathrm{e}^{-D_h t}$ is the well defined exponential function of a matrix, we interpret it at the same time with the semigroup generated by the sectorial operator $D_h$ with domain $V_h^1$. This allows us to use various techniques to derive estimates for the solutions depending on the sectorial properties of $D_h$. Those properties do not depend on the exact construction of $V_h$, i.e. the choice and number of basis functions, see Lemma 5.2.5. We thus derive uniform bounds which must hold for any discretisation $V_h$.

**Lemma 5.2.6.** *Let $\gamma \in [0, 1[$. If Assumption 5.2.4 holds and there is a bound $K$, such that*

$$\|T_h U_0\|_\gamma \leq K \ , \tag{5.25}$$

*then we know that the solution $U : S \to \mathbb{C}^n$ to (5.20) is bounded in the sense that there is a constant $c$, such that*

$$\|T_h U(t)\|_\gamma \leq c \ , \qquad \textit{for all } t \in S \tag{5.26}$$

*and c does not depend on the choice or dimension of $V_h$.*

*Proof.* We use (5.24) and get for any $t \in S$

$$\|T_h U(t)\|_\gamma \leq \|T_h \mathrm{e}^{-tD_h} U(0)\|_\gamma + \|T_h \int_0^t \mathrm{e}^{-(t-s)D_h} P_h N(U(s), s)\,\mathrm{d}s\|_\gamma \ . \tag{5.27}$$

The norms $\|\cdot\|_{h,-1}$ and $\|\cdot\|_{h,1}$ are by definition, see (5.15)-(5.17), connected with the operator $D_h$. Using (5.23), we can switch in (5.27) to these norms. This allows us to use standard semigroup techniques to conclude,

$$\begin{aligned}
\|T_h U(t)\|_\gamma &\leq c\|\mathrm{e}^{-tD_h} U(0)\|_{h,-1}^{\frac{1-\gamma}{2}} \ \|\mathrm{e}^{-tD_h} U(0)\|_{h,1}^{\frac{1+\gamma}{2}} \\
&\quad + c\int_0^t \|\mathrm{e}^{-(t-s)D_h} P_h N(T_h U(s), s)\|_{h,-1}^{\frac{1-\gamma}{2}} \ \|\mathrm{e}^{-(t-s)D_h} P_h N(T_h U(s), s)\|_{h,1}^{\frac{1+\gamma}{2}}\,\mathrm{d}s \ , \\
&\leq c\|U(0)\|_{h,1} + c\int_0^t (t-s)^{-\frac{1+\gamma}{2}} \|P_h N(T_h U(s), s)\|_{h,-1}\,\mathrm{d}s \ .
\end{aligned}$$

We employ Assumption 5.2.4 to get the inequality

$$\|T_h U(t)\|_\gamma \leq c(\|U(0)\|_{h,1} + t^{\frac{1-\gamma}{2}}) + c\int_0^t (t-s)^{-\frac{1+\gamma}{2}} \|T_h U(s)\|_\gamma\,\mathrm{d}s \ ,$$

which allows us to use Gronwall's lemma. This then results by the boundedness of the time interval $S$, into the existence of a uniform bound to $\|T_h U(t)\|_\gamma$ for all times $t \in S$. The constants which appeared in the estimates along the way depended on the sectorial properties of $D_h$. These are independent of the choice and dimension of $V_h$, see Lemma 5.2.5. Hence, as asserted in this lemma, the quality of the final bound in (5.26) does not depend on them as well. □

The bound for the numerical solution, which we derived in the last lemma, helps us to treat the locally Lipschitz continuity of $N$ as Lipschitz, i.e. drop the *locally* in *locally Lipschitz continuous*. This gives rise to better regularity of the discrete solution.

**Lemma 5.2.7.** *Define the bounded time interval $S := [0, T_0]$. Let $U : S \to \mathbb{C}^n$ be the solution to ODE* (5.20)*, and let Assumptions 5.2.1 to 5.2.4 hold. If there exists a bound $K$ such that*

$$\|U(0)\|_{h,1} \leq K \tag{5.28}$$

*then $T_h U$ is Hölder continuous in $V^\theta$ for any $\theta \in [-1, 1[$ in the sense that for all $t_1, t_2 \in S$*

$$\|T_h U(t_1) - T_h U(t_2)\|_\theta \leq C\ |t_1 - t_2|^{\frac{1-\theta}{2}} \ .$$

*We even have for* $\sigma := \min(\alpha_N, \frac{1-\gamma}{2})$ *the time regularity of*

$$U \in \mathrm{C}^1(S, V_h^{-1}) \cap \mathrm{C}(S, V_h^1) \ , \tag{5.29}$$

$$T_h \partial_t U \in \mathrm{C}(S, V^{-1}) \ , \tag{5.30}$$

$$\{t \mapsto P_h N(T_h U(t), t)\} \in \mathrm{C}^\sigma(S, V_h^{-1}) \ . \tag{5.31}$$

*The bounds in these spaces, as well as* $C$*, depend on* $N$ *and the sectorial properties of* $A$*, but not on the exact choice or dimension of* $V_h$*.*

*Proof.* We use the fact that the solution $U$ must satisfy the variation of constants formula. For $0 < t_1 < t_2 < T_0$ this gives rise to

$$\begin{aligned}
\|T_h U(t_2) - T_h U(t_1)\|_\theta \leq{}& \|T_h \Big(\mathrm{e}^{-D_h t_2} - \mathrm{e}^{-D_h t_1}\Big) U(0)\|_\theta \\
&+ \|T_h \int_0^{t_1} \Big(\mathrm{e}^{-D_h(t_2-s)} - \mathrm{e}^{-D_h(t_1-s)}\Big) P_h N(T_h U(s), s)\,\mathrm{d}s\|_\theta \\
&+ \|T_h \int_{t_1}^{t_2} \mathrm{e}^{-D_h(t_2-s)} P_h N(T_h U(s), s)\,\mathrm{d}s\|_\theta \\
\leq{}& I + II + III \ .
\end{aligned}$$

We estimate the three terms separately, using standard semigroup techniques for the semigroup $\mathrm{e}^{-D_h t} : \mathbb{C}^n \to \mathbb{C}^n$. Note that the following estimates use integrals which only exist if $\theta < 1$ and so this is a necessary bound for $\theta$. With the help of (5.23) we can estimate the first term, knowing that $-D_h$ generates a contractive semigroup, by

$$\begin{aligned}
I &= \|T_h \Big(\mathrm{e}^{-D_h(t_2-t_1)} - \mathbb{1}\Big) \mathrm{e}^{-D_h t_1} U(0)\|_\theta \leq \|\cdot\|_{h,-1}^{\frac{1-\theta}{2}} \ \|\cdot\|_{h,1}^{\frac{1+\theta}{2}} \ , \\
&\leq c \big\|(\mathrm{e}^{-D_h(t_2-t_1)} - \mathbb{1}) \mathrm{e}^{-D_h t_1} \ U(0)\big\|_{h,-1}^{\frac{1-\theta}{2}} \ \big\|U(0)\big\|_{h,1}^{\frac{1+\theta}{2}} \ , \\
&\leq c \big\| \int_0^{t_2-t_1} -D_h \mathrm{e}^{-D_h s} \ \mathrm{e}^{-D_h t_1} \ U(0)\,\mathrm{d}s\big\|_{h,-1}^{\frac{1-\theta}{2}} \ \big\|U(0)\big\|_{h,1}^{\frac{1+\theta}{2}} \ , \\
&\leq c \int_0^{t_2-t_1} \big\|\mathrm{e}^{-D_h t_1} \ D_h \ U(0)\big\|_{h,-1}^{\frac{1-\theta}{2}} \,\mathrm{d}s \ \big\|U(0)\big\|_{h,1}^{\frac{1+\theta}{2}} \leq c(t_2 - t_1)\|U(0)\|_{h,1} \ .
\end{aligned}$$

Note that even though the generic constant $c$ might change from step to step, it only depends on the sectorial properties of $D_h$, which themselves depend on the sectorial properties of $A$, see Lemma 5.2.5.

The second and third term can be estimated by additionally exploiting the boundedness of $P_h N(T_h U(t), t)$ in the $V_h^{-1}$-norm. This is true according to Assumption 5.2.4, Lemma 5.2.6 and the simple inequality $\|P_h v\|_{h,-1} \leq \|v\|_{-1}$. This uniform inequality for $P_h$ is true for any $v \in V^{-1}$ and can easily be deduced by the definition of $P_h$ in (5.13).

Using these ingredients we get

$$\begin{aligned} II &\leq \int_0^{t_1} \int_{t_1-s}^{t_2-s} \|T_h D_h \mathrm{e}^{-D_h r} P_h N(T_h U(s), s)\|_\theta \,\mathrm{d}r \,\mathrm{d}s \leq c \int\int \|\cdot\|_{h,-1}^{\frac{1-\theta}{2}} \| \cdot \|_{h,1}^{\frac{1+\theta}{2}} \,\mathrm{d}r \,\mathrm{d}s \\ &\leq c \int_0^{t_1} \int_{t_1-s}^{t_2-s} r^{-\frac{1-\theta}{2}} r^{-2\frac{1+\theta}{2}} \|P_h N(T_h U(s), s)\|_{h,-1} \,\mathrm{d}r \,\mathrm{d}s \\ &\leq c \int_0^{t_1} \int_{t_1-s}^{t_2-s} r^{-\frac{3+\theta}{4}} \,\mathrm{d}r \,\mathrm{d}s \leq c(t_2-t_1)^{\frac{5+\theta}{4}} \ . \end{aligned}$$

We use the uniform bound of $\|P_h N(T_h U(t), t)\|_{h,-1}$ once more and get for the third term

$$\begin{aligned} III &\leq \int_{t_1}^{t_2} \left\|T_h \mathrm{e}^{-D_h(t_2-s)} P_h N(T_h U(s), s)\right\|_\theta \,\mathrm{d}s \leq c \int_{t_1}^{t_2} \|\cdot\|_{h,-1}^{\frac{1-\theta}{2}} \| \cdot \|_{h,1}^{\frac{1+\theta}{2}} \,\mathrm{d}s \ , \\ &\leq c \int_{t_1}^{t_2} (t_2-s)^{-\frac{1+\theta}{2}} \,\mathrm{d}s \leq c(t_2-t_1)^{\frac{1-\theta}{2}} \ . \end{aligned}$$

These three estimates combined, result in a constant $c$ independent of the choice of the subspace $V_h$, such that

$$\|T_h U(t_2) - T_h U(t_1)\|_\theta \leq c|t_1 - t_2|^{\frac{1-\theta}{2}} \ . \tag{5.32}$$

This proves the first assertion in this lemma.

We use the boundedness of $T_h U$ in $V^\gamma$ and deduce from the Hölder continuity of $T_h U$ in (5.32) and the locally Lipschitz continuity of $N$ in Assumption 5.2.4, that $f(t) := P_h N(U(t), t) \in \mathrm{C}^\sigma(S, V_h^{-1})$ with $\sigma = \min(\alpha_N, \frac{1-\theta}{2})$. The norm of $f$ in $\mathrm{C}^\sigma(S, V_h^{-1})$ is bounded from above independent of the choice of $V_h$. This is the proof of (5.31).

We can then apply known regularity results for non-homogeneous equations of the form $\partial_t u = -D_h u + f(t)$, see [28, thm.4.3.1]. Exploiting the regularity of the initial value, see (5.28), we then know that

$$U \in \mathrm{C}^1(S, V_h^{-1}) \cap \mathrm{C}(S, V_h^1) \ , \tag{5.33}$$

with norm bounded from above depending on the sectorial properties of $D_h$, but independent of the choice of $V_h$. This gives the desired quality (5.29). Consequently, $\partial_t U$ is bounded and continuous in $V_h^{-1}$, which gives by (5.18) the continuity and boundedness of $T_h \partial_t U$ in $V_h^{-1}$. □

With the previous lemmata we can derive the convergence of approximations $U_j$, where each $U_j$ is the solution to a discretisation of the PDE (5.19) on a space $V_{h_j}$ in the form of (5.20). The sequence of spaces $V_{h_j}$ fills up the whole space in the limit for $j \to \infty$. The subscript $h_j$ shall suggest that this is usually accomplished by a space discretisation with a grid size $h_j \to 0$.

**Theorem 5.2.8.** *Assume that the mapping $N : V^\gamma \times \mathbb{R} \to V^{-1}$ is locally Lipschitz in the first argument and locally Hölder continuous in the second, as described in Assumption 5.2.4. We consider a sequence of finite dimensional subspaces $V_{h_j} \subset V$ of the form*

(5.10), *which in the limit are dense in* $V$. *By this we mean*

$$\begin{aligned} &\text{for all } v \in V \text{ there exists a sequence } v_j \in V_{h_j}^1, j = 1, 2, \cdots \text{ , such that,} \\ &T_{h_j} v_j \to v \text{ in } V \ . \end{aligned} \tag{5.34}$$

*Furthermore, we need for the sequence* $V_{h_j}$ *that the respective projections* $P_{h_j}$ *are stable, as described in Assumption 5.2.3. Additionally let Assumptions 5.2.1-5.2.2 hold.*

*For each* $j$ *we discretise the operator differential equation* (5.19) *to get an ODE of the form* (5.20) *on a finite time interval* $S := [0, T_0]$. *Let* $U_j(0) = Q_{h_j} u_0 \in V_{h_j}$ *be the sequence of initial data for the discretised problems. We assume that* $u_0 \in V^1$ *and that there is a constant* $K$ *such that*

$$T_{h_j} Q_{h_j} u_0 \to u_0 \ , \quad \text{in } V^{-1} \ , \tag{5.35}$$

$$\|Q_{h_j} u_0\|_{h_j,1} \leq K \ . \tag{5.36}$$

*The functions* $U_j : S \to \mathbb{C}^{N_j}$, *which are the solutions to the set of ODE's, then converge along a subsequence to a solution* $u$ *of the PDE* (5.19), *in the sense that for all* $\theta \in [-1, 1[$,

$$T_{h_j} U_j \xrightarrow{j \to \infty} u \qquad \text{in } \mathrm{C}(S, V^\theta) \ . \tag{5.37}$$

*Proof.* Since the proof is quite long, we give a sketch of the proof in a table of contents manner first.

(i) gather results of previous lemmata

(ii) show convergence of $T_{h_j} U_j$ to some $u_\infty$ in $\mathrm{C}(S, V^\theta)$ for any $\theta \in [-1, 1[$

(iii) deduce weak $\mathrm{L}^2(S, V^{-1})$ convergence of each term in the discretised formulation

(iii.a) deduce the convergence $\partial_t T_{h_j} U_j \to \partial_t u_\infty$

(iii.b) deduce the convergence $\{t \mapsto T_{h_j} P_{h_j} N(U_j(t), t)\} \to \{t \mapsto N(u_\infty(t), t)\}$

(iii.c) deduce the convergence $T_{h_j} D_{h_j} U_j \to A u_\infty$

(iv) final consequence, $u_\infty$ is a solution to (5.19)

(i) *gather results of previous lemmata.* Our assumptions allow us to apply Lemma 5.2.7. Then for $f(t) := T_{h_j} P_{h_j} N(U_j(t), t)$ and any $\theta \in [-1, 1]$ there is a $\delta > 0$ and an upper bound $K$, independent of $j$, such that

$$\|T_{h_j} U_j\|_{\mathrm{C}^{\frac{1-\theta}{2}}(S, V^\theta)} \leq K \ , \tag{5.38}$$

$$\|f(t)\|_{\mathrm{C}^\delta(S, V^{-1})} \leq K \ , \tag{5.39}$$

$$\|T_h \partial_t U_j(t)\|_{\mathrm{C}(S, V^{-1})} \leq K \ . \tag{5.40}$$

(ii) *show convergence of* $T_{h_j} U_j$ *to some* $u_\infty$ *in* $\mathrm{C}(S, V^\theta)$. Take any $\theta \in [-1, 1[$. By the uniform bound of $T_{h_j} U_j$ in $V^1$, see (5.38), and the assumed compact embedding, see

(5.9), we deduce that for any $t \in S$ the sequence $\left\{T_{h_j} U_j(t)\right\}_j$ is precompact in $V^\theta$. Additionally, the family of functions $T_{h_j} U_j : S \to V^\theta$ is equicontinuous as a consequence of the equi Hölder continuity in (5.38). This allows us to employ a variant of the Arzela Ascoli theorem for families of functions with values in general Banach spaces, see [23, p. 233]. Hence, $T_{h_j} U_j$ must converge in $\mathrm{C}(S, V^\theta)$ (along a not relabelled subsequence)

$$T_{h_j} U_j \to u_\infty \qquad \text{in} \quad \mathrm{C}(S, V^\theta) \ . \tag{5.41}$$

(iii) *deduce suitable convergence of each term in their discretised formulation .*
We are left to prove, that $u_\infty$ is a solution to the original problem (5.19). For this we show that the terms in the discretised version (5.20) converge piecewise to their corresponding counterparts in (5.19). To be precise, we have to show that their images under the mapping $T_h$ converge in $V^{-1}$. In order to connect this to the convergence in (5.41), we consider the subsequence of $T_{h_j} U_j$, such that (5.41) holds. We do so without relabelling.

(iii.a) *deduce the convergence* $\partial_t T_{h_j} U_j \to \partial_t u_\infty$ *in* $\mathrm{L}^2(S, V^{-1})$
By (5.40) we have the uniform boundedness of $\partial_t T_{h_j} U_j(t) = T_{h_j} \partial_t U_j(t)$ in $V^{-1}$ for all $t$ and $j$. Together with (5.38) this gives that the function $T_{h_j} U_j$ is bounded in $\mathrm{H}^1(S, V^{-1})$ uniformly in $j$. Thus, a subsequence must converge weakly to some $\tilde{u} \in \mathrm{H}^1(S, V^{-1})$. This $\tilde{u}$ must then coincide with the limit $u_\infty$ in (5.41), which means that the weak limit $\tilde{u}$ is unique. Therefore, the whole sequence $\partial_t T_{h_j} U_j(t)$ converges weakly to $\partial_t u \in \mathrm{L}^2(S, V^{-1})$.

(iii.b) *deduce the convergence* $\{t \mapsto T_{h_j} P_{h_j} N(U_j(t), t)\} \to \{t \mapsto N(u_\infty(t), t)\}$.
By the convergence in (5.41) and the continuity of $N$, see Assumption 5.2.4, we have

$$N(T_{h_j} U_j(t), t) \to N(u(t), t) \qquad \text{in } \mathrm{C}(S, V^{-1}) \ . \tag{5.42}$$

From this we will deduce the weak convergence of $T_{h_j} P_{h_j} N(T_{h_j} U_j(t), t)$ to $N(u(t), t)$ in $V^{-1}$ for all fixed $t \in S$. First we show a boundedness in $V^{-1}$ as follows,

$$\begin{aligned} \|T_{h_j} P_{h_j} N(U_j(t), t)\|_{-1} &= \sup_{v \in V} \frac{\left(T_{h_j} P_{h_j} N(U_j(t), t), v\right)_X}{\|v\|_1} = \sup_{v \in V} \frac{\left(N(U_j(t), t), P_{h_j} v\right)_X}{\|v\|_1} \ , \\ &\leq \sup_{v \in V} \|N(U_j(t), t)\|_{-1} \frac{\|P_{h_j} v\|_1}{\|v\|_1} \ . \end{aligned}$$

The assumption (5.14) gives the boundedness of the quotient $\|P_{h_j} v\|_1 / \|v\|_1$. Employing (5.42), we deduce the uniform boundedness of $\|T_{h_j} P_{h_j} N(U_j(t), t)\|_{-1}$ for all $j$. The space $V^{-1}$ is a Hilbert space and so this boundedness results in the weak convergence of a subsequence.

We analyse for fixed $t \in S$ the weak limit of $N_j(t) := T_{h_j} P_{h_j} N(U_j(t), t)$ (in a not relabelled subsequence) by the following estimate. Take an arbitrary element $v \in V$ and choose a recovery sequence $v_j$ as in (5.34). We use the identity

$$\langle T_{h_j} P_{h_j} N(T_{h_j} U_j(t), t), T_{h_j} v_j \rangle = \langle N(T_{h_j} U_j(t), t), T_{h_j} v_j \rangle$$

to get

$$\begin{aligned}|\langle N_j(t),v\rangle - \langle N,v\rangle| &= |\langle N_j(t),v\rangle \pm \langle N_j(t),T_{h_j}v_j\rangle \pm \langle N,T_{h_j}v_j\rangle - \langle N,v\rangle| \ , \\ &\le |\langle N_j(t),v-T_{h_j}v_j\rangle| + |\langle N(T_{h_j}U_j(t),t)-N(u(t),t),T_{h_j}v_j\rangle| \\ &\quad + |\langle N(u(t),t),v-T_{h_j}v_j\rangle| \ , \\ &\le \big(\|N_j(t)\|_{-1} + \|N(u(t),t)\|_{-1}\big)\|v-T_{h_j}v_j\|_1 \\ &\quad + \|N(U_j(t),t)-N(u(t),t)\|_{-1}\ \|T_{h_j}v_j\|_1 \ .\end{aligned}$$

The first addend vanishes for $j\to\infty$ by the boundedness of $N(u(t),t)$ and $N_j(t)$ and the convergence of $T_{h_j}v_j$. The second addend vanishes by the convergence in (5.41). Thus, we have the weak convergence $T_{h_j}P_{h_j}N(U_j(t),t) \rightharpoonup N(u(t),t)$. By the uniqueness of the limit we deduce that the whole subsequence must converge weakly to this limit for any fixed $t$. In combination with the boundedness of $T_{h_j}P_{h_j}N(U_j(t),t)$ on $V^{-1}$, which is uniform for all $t$, this results in the weak convergence

$$\big\{t\to T_{h_j}P_{h_j}N(U_j(t),t)\big\} \ \rightharpoonup\ \big\{t\to N(u(t),t)\big\} \qquad \text{in } \mathrm{L}^2(S,V^{-1}) \ . \tag{5.43}$$

(iii.c) *deduce convergence $T_{h_j}D_{h_j}U_j \to Au_\infty$.* We have already shown the weak convergence of the discrete time derivative and of the nonlinear term. This immediately gives, by equality (5.20), that there is a $w\in V^{-1}$ such that

$$\partial_t T_{h_j}U_j - T_hP_hN(T_{h_j}U_j(t),t) = T_{h_j}D_{h_j}U_j \rightharpoonup w \ , \qquad \text{in } \mathrm{L}(S,V^{-1}) \ . \tag{5.44}$$

We are left to show that $w=Au$.

By the uniform (in $j$ and $t$) boundedness of $\|T_{h_j}U_j(t)\|_1$ we know that along a subsequence $T_{h_j}U_j$ must converge weakly in $\mathrm{L}^2(S,V^1)$. Consequently, $AT_{h_j}U_j$ must converge weakly in $\mathrm{L}^2(S,V^{-1})$. We show that this weak limit must coincide with $w$. For this purpose we investigate the following. Take any $v\in V^1$, $t_1<t_2\in S$, consider $v_j\in V^1_{h_j}$ to be a recovery sequence for $v$ as in (5.34). We derive the estimate

$$\begin{aligned}\Big|\int_{t_1}^{t_2}\langle T_{h_j}D_{h_j}U_j(t) - AT_{h_j}U_j(t)\ ,v\rangle\,\mathrm{d}t\Big| &\le \int_{t_1}^{t_2}|\langle T_{h_j}D_{h_j}U_j(t)-AT_{h_j}U_j(t)\ ,T_{h_j}v_j\rangle| \\ &\quad + |\langle T_{h_j}D_{h_j}U_j(t)-AT_{h_j}U_j(t)\ ,v-T_{h_j}v_j\rangle|\,\mathrm{d}t \ .\end{aligned}$$

Due to the construction of $D_{h_j}$ we know, when acting on functions in the image of $T_h$, that $T_{h_j}D_{h_j}U_j(t)$ and $AT_{h_j}U_j(t)$ coincide. Therefore, the first term vanishes and we get

$$\begin{aligned}&\Big|\int_{t_1}^{t_2}\langle T_{h_j}D_{h_j}U_j(t) - AT_{h_j}U_j(t)\ ,v\rangle\,\mathrm{d}t\Big| \\ &\qquad\le \int_{t_1}^{t_2}\|T_{h_j}D_{h_j}U_j(t)-AT_{h_j}U_j(t)\|_{-1}\ \|v-T_{h_j}v_j\|_1\,\mathrm{d}t \ .\end{aligned} \tag{5.45}$$

Having shown the boundedness of the left hand side of (5.44) uniformly in $j$ and $t$, we

know that $\|T_{h_j} D_{h_j} U_j(t)\|_{-1}$ is also uniformly bounded. Furthermore, by (5.38), the term $\|AT_{h_j} U_j(t)\|_{-1} = \|U_j(t)\|_{h,1}$ stays bounded as well. Thus, we deduce from (5.45) with the help of (5.34),

$$|\int_{t_1}^{t_2} \langle T_{h_j} D_{h_j} U_j(t) - AT_{h_j} U_j(t)\ , v\rangle \,\mathrm{d}t| \qquad \leq c \int_{t_1}^{t_2} \|v - T_{h_j} v_j\|_1 \,\mathrm{d}t \to 0\ . \tag{5.46}$$

Consequently, for any function $v : S \to V^1$, which is piecewise constant in time, we get the equality

$$\int_S \langle w(t), v(t)\rangle \,\mathrm{d}t = \lim_{j\to\infty} \int_S \langle T_{h_j} D_{h_j} U_j(t), v(t)\rangle \,\mathrm{d}t = \lim_{j\to\infty} \int_S \langle AT_{h_j} U_j(t)\ , v(t)\rangle \,\mathrm{d}t\ . \tag{5.47}$$

Since those piecewise constant functions are dense in $\mathrm{L}^2(S, V^{-1})$, we have by [20, Thm 8.40] that $AT_{h_j} U_j$ converges weakly in $\mathrm{L}(S, V^{-1})$ to $w$ as well.

By Assumption 5.2.1, $A$ is a weakly closed operator, and thus there holds

$$Au_\infty = w = (\text{weak}) \lim_{j\to\infty} AT_{h_j} U_j = (\text{weak}) \lim_{j\to\infty} T_{h_j} D_{h_j} U_j \qquad \text{in } \mathrm{L}^2(S, V^{-1})\ . \tag{5.48}$$

Note that this limit is again unique and hence, the whole sequence must converge weakly.

(iv) *final consequence, $u_\infty$ is a solution to* (5.19). We have shown that all discrete terms in (5.20) converge (in $\mathrm{L}^2(S, V^{-1})$) to their continuous counterpart in (5.19) along a subsequence as chosen in part (ii) of the proof. Thus we get the equality in $\mathrm{L}^2(S, V^{-1})$,

$$\partial_t u_\infty(t) = -Au_\infty(t) + N(u_\infty(t), t)\ . \tag{5.49}$$

By the convergence (5.35) we also know that $u_\infty(0) = u_0$. The limit function $u_\infty : S \to V^1$ is thus a solution to (5.19). This proves the assertion of this theorem. Note that by (5.35) and (5.36) and the compact embedding of the interpolation spaces, see (5.9), $T_h U_j(0)$ converges in all spaces $V^\theta$, $\theta \in [-1, 1[$ to $u_0$. In $V^1$ it still converges weakly. But it suffices to demand only (5.35) and (5.36), as this implies convergence in all $V^\theta$. □

We can now state the convergence of numerical solutions to the problem (5.1) as an application of the Theorem 5.2.8. This is done in two steps. In a following theorem we prove the convergence of the semi-discretised problem. The convergence of the solutions to the fully discretised problem is then a simple corollary. In order to apply the above convergence results, we modify our problem, such that it satisfies Assumption 5.2.4. We define for a positive number $K > 0$ the cut-off function

$$\zeta_K(s) := \begin{cases} -K\ , & \text{if } s < -K\ . \\ s\ , & \text{if } s \in [-K, K]\ . \\ K\ , & \text{if } s > K\ . \end{cases} \tag{5.50}$$

We substitute in (5.1) $\zeta_K(\mathbb{L})$ for $\mathbb{L}$ defined in (3.8). When choosing $K$ large enough, this does not alter the problem, if we can guarantee that $\mathbb{L}(u(t))$ stays bounded for the

original solution $u$ at all times $t$. For this boundedness we need the quality of the given initial data $u_0 \in \mathrm{H}^1(\Omega)$ and locally Hölder continuous data $p \in \mathrm{C}^\delta([0,T_0])$, $\delta > \frac{1}{2}$, such that $\ell(t) := \int_\Omega x u_0(x)\,\mathrm{d}x + \int_0^t p(t) \in ]0,1[$ for all $t \in [0,T_0]$. From the analytic treatment in Section 4.2 we then know that there exists a solution $u$ of (3.9) in the sense of (5.1), and it is bounded in $\mathrm{H}^1(\Omega)$ for all times. Since $\mathbb{L}$ is a linear form in $\mathrm{H}^{\frac{1}{2}+\varepsilon}$, this gives the needed guarantee.

The modified problem formulation reads

$$\forall t \in S\ ,\ \forall v \in H^1(\Omega) \tag{5.51}$$

$$\begin{aligned}\Big(\partial_t u(t), v\Big)_{\mathrm{L}^2} = &-\frac{\nu^2}{\tau}\Big(\partial_x u(t), \partial_x v(t)\Big)_{\mathrm{L}^2} \\ &- \frac{1}{\tau}\Big([\psi' - \tau p(t) - \zeta_K\Big(\mathbb{L}(u(t))\Big)]\zeta_K(u(t))\ ,\ \partial_x v(t)\Big)_{\mathrm{L}^2}\ .\end{aligned} \tag{5.52}$$

Note that the solution is also unique, since it fits in the same framework of Theorem 4.1.2. We carry this over to the discretisation which reads in the semi discretised form

$$M_h\dot{U}(t) = -\frac{\nu^2}{\tau}A_hU(t) - \frac{1}{\tau}\widehat{B}U(t) + p(t)BU(t) + \zeta_K\Big((L\ ,\ U(t))_{\mathrm{e}}\Big)B\begin{pmatrix}\zeta_K(U_1(t))\\ \zeta_K(U_2(t))\\ \vdots\end{pmatrix}\ . \tag{5.53}$$

As initial value, $U$ must satisfy $U(0) = Q_h(u_0)$ for a suitable mapping $Q : \mathrm{H}^1(\Omega) \to V_h$. Note that $U(t)$ can also be interpreted as a function $\sum_{j=1}^N U_j(t)\varphi_j \in \mathrm{H}^1(\Omega)$. As a first remark concerning these two interpolations, we would like to point out, that $U$ fulfils (5.51) when tested with functions from $V_h$. This helps us to carry over some properties from the exact to the semi-discrete problem as stated in the next lemma.

**Lemma 5.2.9.** *If $v_1(x) := 1 \in V_h \subset \mathrm{H}^1(\Omega)$ then the solution $U$ to (5.2) or (5.53) satisfies for all times $t > 0$ in the existence interval*

$$\int_\Omega T_hU(t,x)\,dx = \int_\Omega T_hU(0,x)\,dx\ . \tag{5.54}$$

*Furthermore if $v_2(x) := x \in V_h \subset \mathrm{H}^1(\Omega)$ and $\int_\Omega U(0,x)\,dx = 1$, then the solution $U$ to (5.2) satisfies*

$$\int_\Omega xT_hU(t,x)\,dx = \int_\Omega xT_hU(0,x)\,dx + \int_0^t p(t)\,dt\ , \tag{5.55}$$

*for all times $t > 0$, such that the solution exists up to this time. Note that the last assertion is then also true for solutions to (5.53), if $\zeta_K(\mathbb{L}(U(s))) = \mathbb{L}(U(s))$ and $\zeta_K(U_j(s)) = U_j(s)$ for all $s \in [0,t]$. In this case, the modified problem is equivalent to the unmodified version on this time interval.*

*Proof.* The first ingredient of the proof is the fact that by the duality of interpreting $v_1(t), v_2(t) \in \mathbb{R}^N$ as vectors in $\mathbb{R}^N$, or respectively as functions in $V_h \subset \mathrm{H}^1(\Omega)$, we have

for any $v \in V_h$ the standard identities

$$\Big(v_1, M_h v_2\Big)_{\mathrm{e}} = \Big(v_1, v_2\Big)_{\mathrm{L}^2} , \qquad \Big(v_1, A_h v_2\Big)_{\mathrm{e}} = \Big(\partial_x v_1, \partial_x v_2(t)\Big)_{\mathrm{L}^2}^2 .$$

Moreover, we have

$$\Big([\widehat{B} - \tau p(t) B - \zeta_K \Big((L\ ,\ v_1)_{\mathrm{e}}\Big) B] v_1, v_2\Big)_e = \Big([\psi' - \tau p(t) - \zeta_K \Big(\mathbb{L}(v_1)\Big)] v_1, \partial_x v_2(t)\Big)_{\mathrm{L}^2} ,$$

due to the construction of $A_h$, $M_h$, $B$, $\widehat{B}$ and $L$. This equality allows us to use $U$, the solution to (5.2), in place of $u$ in (5.1) and test with $v \equiv 1 \in V_h$ to get

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_\Omega U(t,x)\,\mathrm{d}x = 0.$$

The same reasoning works for solutions to (5.53) and hence proves the first stated equality in the lemma.

For the second equality in this lemma we again replace in (5.1) $u$ by $U$ and test with $v_2(x) = x \in V_h$. Then, employing the exact formula for $\mathbb{L}$ results in the second inequality. This shows that the choice for $\mathbb{L}$, see (3.8), is also the correct choice in order to control the first moment of solutions in the semi-discretised problem. □

We still need to define the choice of the mapping which gives for an $u_0 \in \mathrm{H}^1(\Omega)$ a corresponding discrete initial value $U_h(0)$. Say that we have an underlying grid on the interval $[0,1]$ and our discretisation consists of piecewise affine functions on this grid. We give here two possible choices together with the needed quality of $u_0$. The reason why we give two choices is, that in our numerical experiments we used the first version, while the second version needs less regularity for $u_0$, and therefore deserves to be mentioned as well. The first possibility is

$$U_h(0) = c_* I_h u_0 , \qquad u_0 \in \mathrm{C}^1([0,1]), \qquad \int_\Omega u_0(x)\,\mathrm{d}x = 1 , \tag{5.56}$$

with $I_h$ being the nodal interpolant and $c_* = 1/\int_\Omega (T_h I_h u_0)(x)\,\mathrm{d}x$. The choice of $c_*$ guarantees that $T_h U_h(0)$ has also mean value of one. Another possible choice is

$$U_h(0) = P_h u_0 , \qquad u_0 \in \mathrm{H}^1(\Omega), \qquad \int_\Omega u_0(x)\,\mathrm{d}x = 1 , \tag{5.57}$$

where $P_h$ is the $\mathrm{L}^2$-projection on the space of all piecewise affine functions having a mean value of one.

**Theorem 5.2.10.** *Set $h_j := 1/j$ for $j = 1, 2, \cdots$. Consider $\mathcal{P}^1_{h_j}$ to be the space of continuous and piecewise linear functions on an equidistant partition of $\Omega = [0,1]$ into intervals of length $h_j$.*

*Let the assumptions of the existence Theorem 4.2.5 and 4.1.2 hold, such that there exists a unique solution to the PDE (3.9) on the finite time interval $S = [0, T_0]$. Fur-*

*thermore, let $U_j$ be the solution of the ODE* (5.2)*, which is the (semi)discretised version of the PDE on the space $\mathcal{P}^1_{h_j}$. For the discrete initial value $U_j(0)$ we choose for all $j$ either* (5.56) *or for all $j$* (5.57)*. The necessary quality of $u_0$, as imposed in* (5.56) *or* (5.57)*, shall be satisfied.*

*Then there is a critical value $j_*$, such that for all $j > j_*$ the solution $U_j$ exists on the whole time interval $S$ and it converges to $u$ in the sense that for any $\theta \in [-1, 1[$*

$$T_{h_j} U_j \xrightarrow{j\to\infty} u \qquad \text{in } \mathrm{C}(S, \mathrm{H}^\theta_m(\Omega)) \ . \tag{5.58}$$

*Proof.* We aim at showing that the PDE (3.9) or its weak formulation (5.1) and its discretisation (5.2) fit into the framework of Theorem 5.2.8. For this reason we have to modify the PDE (3.9) into an equivalent one. According to the assumptions, we have constants $\nu, \tau > 0$ and $\psi \in \mathrm{H}^1(\Omega)$ and also $p \in \mathrm{C}^{\alpha_N}(S, \mathbb{R})$ for an $\alpha_N > \frac{1}{2}$ as the given data of the PDE (3.9).

We define the complex Hilbert spaces $V^1$ and $X$ and the operator $A$ as

$$\begin{aligned} V^1 &:= \mathrm{H}^1_m(\Omega) := \Big\{ v \in \mathrm{H}^1(\Omega) : \int_\Omega v \,\mathrm{d}x = 1 \Big\} \ , \quad (v_1, v_2)_1 := \frac{\nu^2}{\tau} \int_\Omega \partial_x v_1(x)\ \partial_x \overline{v_2}(x) \,\mathrm{d}x \\ X &:= \mathrm{L}^2(\Omega) \\ A &: V^1 \subset V^{-1} \to V^{-1} \ , \qquad\qquad \langle Au, v\rangle = \frac{\nu^2}{\tau} \int_\Omega \partial_x u \partial_x v \,\mathrm{d}x \ . \end{aligned}$$

The space $V^{-1}$ is the usual dual of $V^1$. The nonlinearity in the PDE is considered by defining the mappings $N, N_0 : \mathrm{C}(\Omega) \to \mathrm{H}^{-1}(\Omega)$, such that for all $v \in \mathrm{C}(\Omega)$ :

$$\begin{aligned} \langle N_0(v,t), w\rangle &:= \Big( [\psi' - \tau p(t) - \zeta_K\Big(\mathbb{L}(v)\Big)] \zeta_K(v), \partial_x w(t) \Big)_{\mathrm{L}^2} \quad \text{for all } w \in \mathrm{H}^1(\Omega) \ , \\ N(v,t) &:= N_0(v+1, t) \ . \end{aligned} \tag{5.59}$$

The function $\zeta_K$ is the cut-off function as described in (5.50). According to Corollary 4.1.4, the solution $u$ is in the space $\mathrm{C}(S, \mathrm{H}^1_m(\Omega))$. We define the cut-off barrier to be $K := 1 + \max_{t\in S}\{|\mathbb{L}u(t)|\} + \max_{t\in S}\{|u(t)|_{\mathrm{C}}\}$ and deduce that $u$ coincides with the unique solution to the modified version $\partial_t u = -Au + N_0(u,t)$, $u(0) = u_0$. In order to work in the spaces of mean-value free functions we switch to the equivalent problem of seeking $w$, such that

$$\partial_t w = -Aw + N(w,t) \ , \qquad\qquad w(0) = u_0 - 1 \ , \tag{5.60}$$

and clearly $u \equiv w + 1$.

As shown in Lemma 5.2.9, all discrete solution $U_j : S \to \mathbb{C}^{n_j}$ satisfy $\sum_{j=1}^{n_j} (U_j(t))_j \equiv 1$. We also switch in the discretised version from (5.53) to

$$\partial_t W_j = -D_{h_j} W + P_{h_j} N(T_{h_j} W, t) \ , \qquad W_j(0) = Q_{h_j}(u_0 - 1) \ , \tag{5.61}$$

where $D_{h_j}$ is defined as in (5.11). Similar to the case of (5.60), we have the identity

$U_j = W_j + (1, 1, \cdots, 1) \in \mathbb{C}^{n_j}$. As a remark we note that for real valued vectors $U_j(0)$ the solution is also real valued for all times.

Now that we have clarified the setting, we need to show that the assumptions of Theorem 5.2.8 hold. The operator $A$ is the well known Laplacian for mean-value free functions, and thus Assumptions 5.2.1 and 5.2.2 are satisfied. We exploit the modification of the nonlinearity $N$, which we made by inserting the cut-off function $\zeta_K$, to show Assumption 5.2.4. In order to do this, we inspect for $v_1, v_2 \in \mathrm{H}^1_m(\Omega)$ and $s_1, s_2 \in S$ the difference

$$\|N(v_1, s_1) - N(v_2, s_2)\|_{-1} \leq \|N(v_1, s_1) - N(v_1, s_2)\|_{-1} + \|N(v_1, s_2) - N(v_2, s_2)\|_{-1} \ . \tag{5.62}$$

For the first difference we use the assumed Hölder continuity of $p$ to get

$$\|N(v_1, s_1) - N(v_1, s_2)\|_{-1} \leq c\|p\|_{\mathrm{C}^{\alpha_N}} \ |s_1 - s_2|^{\alpha_N} \ (\|v_1\|_{\mathrm{L}^2} + 1) \ . \tag{5.63}$$

For the second term in (5.62) we use that for any $a_1, a_2 \in \mathbb{R}$,

$$\begin{aligned} \Big|\zeta_K(a_1)\zeta_K(v_1) - \zeta_K(a_2)\zeta_K(v_2)\Big| &= \Big|\zeta_K(a_1)(\zeta_K(v_1) - \zeta_K(v_2)) + (\zeta_K(a_1) - \zeta_K(a_2))K\Big| \\ &\leq K(|v_1 - v_2| + |a_1 - a_2|) \end{aligned} \tag{5.64}$$

and so we get for $c_{\mathbb{L}} := \sup\{|\mathbb{L}(v)| : v \in \mathrm{C}(\bar{\Omega})\}$,

$$\|N(v_1, s_2) - N(v_2, s_2)\|_{-1} \leq c\Big(\|\psi'\|_{\mathrm{L}^2} + \tau\|p\|_{\mathrm{L}^\infty} + K + c_{\mathbb{L}}K\Big)\|v_1 - v_2\|_{\mathrm{C}} \ . \tag{5.65}$$

The generic constants $c$ in (5.63) and (5.65) depend on $\nu$ and $\tau$. We know that for $\gamma > \frac{1}{2}$ we have the compact embedding $V^\gamma \hookrightarrow \mathrm{C}(\Omega)$. Thus, combining (5.62)-(5.65) we get the first estimate (5.21) in Assumption 5.2.4. The second inequality in this assumption follows easier. By the boundedness of $\zeta_K$ and $p$ we get for a constant $c = c(\tau, \nu)$,

$$\|N(v, s)\|_{-1} \leq c\Big(\|\psi'\|_{\mathrm{L}^2} + \tau\|p\|_{\mathrm{L}^\infty} + K\Big)\Big(\|v\|_{\mathrm{C}} + 1\Big) \ ,$$

such that by the same embedding (5.22) must be satisfied.

The discrete subspaces $V_{h_j} = \mathcal{P}^1_{h_j} \cap \mathrm{H}^1_m(\Omega)$ consist of piecewise affine and mean-value free functions on an equidistant decomposition of $\Omega = [0, 1]$. It is known that for $j \to \infty$ they are dense in $\mathrm{H}^1_m(\Omega)$ as desired in (5.34). The needed stability of the $\mathrm{L}^2$ projection in (5.14) is satisfied according to Bramble et al., see [4]. Hence, in order to apply Theorem 5.2.8 we are only left to show the quality of the discrete initial values as desired in (5.35) and (5.36).

The choice for $U_{h_j}(0)$ in (5.56) or in (5.57) clearly defines the projection of $Q_h$ in (5.61). We know that for both choices we have the $\mathrm{L}^2$ convergence $T_h U_j(0) \to u_0$. Thus, there must hold for $T_h W_j = -1 + T_h U_j(0)$ and $w_0 = u_0 - 1$ that $T_h W_j \to w_0$ in $\mathrm{L}^2(\Omega)$. This gives the convergence of the initial value as assumed in (5.35). The boundedness of $\|T_h W_{h_j}(0)\|_1 = \|T_h U_{h_j}(0)\|_1$ stems from $u_0$ being in $\mathrm{C}^1(\Omega)$ in the case of (5.56), whereas

in the case of (5.57) it stems from the stability of the $\mathrm{L}^2$-projection.

We have gathered all necessary assumptions to apply Theorem 5.2.8. This gives a convergence of $T_{h_j} W_j \to w$ in $\mathrm{C}(S, V^\theta)$, which is equivalent to $T_{h_j} U_j \to u$ in $\mathrm{C}(S, V^\theta)$. This convergence allows us to drop the use of the cut-off function $\zeta_K$. By the continuity of $\mathbb{L}$ on $V^\gamma$ we get that there is an $j_*$, such that for all $j > j_*$ there holds

$$\sup_{t\in S} |\mathbb{L}(T_h U_j(t))| + \sup_{t\in S} \|T_h U_j(t)\|_{\mathrm{C}} \leq 1 + \sup_{t\in S} |\mathbb{L}(u(t))| + \sup_{t\in S} \|u(t)\|_{\mathrm{C}} =: K \ .$$

Thus, the artificially introduced cut-off function in the nonlinear term $N$ does not alter our problem. □

In a final step we can deduce from the latter theorem that also the fully discretised problem, as described in Section 5.1 and used in experiments in Section 5.3, converges to the exact solution. For this the space and the time discretisation must get finer in a suitable way.

**Corollary 5.2.11.** *We consider a finite time interval $S = [0, T_0]$. Let the assumptions of Theorem 4.1.2 and Theorem 4.2.6 hold, such that a solution to the PDE (3.9) exists on the interval $S$. Let $V_{h_j}$ be a sequence of discretisations of $\mathrm{H}^1(\Omega)$ as described in Theorem 5.2.10. Each $V_{h_j}$ belongs to the spatial grid sizes $h_j$ and $h_j \to 0$ as $j \to \infty$. Furthermore, let the mapping of the initial data $u_0 \in \mathrm{H}^1(\Omega)$ into $V_{h_j}$ to the discrete initial datum be either as in (5.56) or as in (5.57).*

*There exists a sequence $\{k(h_j)\} \subset \mathbb{R}$, such that for all pairs of space and time step sizes $(h_j, k_j)$, such that $0 < k_j \leq k(h_j)$ the discretisation error vanishes. Using the notation*

| | |
|---|---|
| $u$ | *for exact solution of PDE (3.9),* |
| $U_j^m$ | *for the FEM approximation of $u$ at time $t_{m,j} := mk_j$ according to Section 5.1, discretised with spatial grid size $h_j$ and time step size $k_j$ ,* |

*we state that for any $\theta \in [-1, 1[$,*

$$\sup_{m\in\mathbb{N}, 0\leq t_{m,j}\in S} \|T_{h_j} U_j^m - u(t_{m,j})\|_\theta \xrightarrow{j\to 0} 0 \ . \tag{5.66}$$

*As a consequence of the embedding $\mathrm{H}^{\frac{1}{2}+\varepsilon}(\Omega) \hookrightarrow \mathrm{C}(\bar{\Omega})$ this also results in convergence in the $\mathrm{C}(\bar{\Omega})$ norm.*

*Proof.* From Theorem 5.2.10 we know that there exists a sequence of solutions $U_{h_j}$, to the semi-discretised problem (5.2). Each one is an ODE of the form (5.20). These solutions converge in all spaces $\mathrm{C}(S, \mathrm{H}^\theta(\Omega))$, $\theta \in [-1, 1[$ to $u$.

Another consequence of Theorem 5.2.10 is, that there exists $K \in \mathbb{R}$ and an $j_* \in \mathbb{N}$, such that

$$\text{for all } j > j_* : \qquad \sup_{t\in S} \left(\|T_{h_j} U_{h_j}(t)\|_{\mathrm{C}} + |\mathbb{L}(T_{h_j} U_{h_j}(t))|\right) \leq K \ . \tag{5.67}$$

We will use this knowledge in order to transform our original ODE into an equivalent one with Lipschitz right-hand side.

In the spirit of the proof of the Theorem 5.2.10 we modify the nonlinear part of the ODE with the cut-off function $\zeta_{K+1}$, as defined in (5.50). Hence, we can use standard arguments for the numerical integration of ODE's with Lipschitz right-hand side.

We denote by $\{U^m_{h_j,k}\}_m$ the set of solution vectors of the fully discretised problem with time step size $k$. The time discretisation we chose, i.e. the Crank Nicholson scheme, can be interpreted as a Runge-Kutta scheme of second order or a multistep method. In both cases it is an A-stable method of order two, and so we can use known theory to state convergence of the fully discretised problem to solutions of the semi-discretised problems. We refer to a convergence result by Hairer and Wanner, [18, Thm.6.11], which states, assuming the Lipschitz continuity of the right hand side, that for fixed $j$

$$e(h_j,k) := \sup_{m\in\mathbb{N},0\le mk\le T_0} \|T_{h_j}U^m_j - T_{h_j}U_{h_j}(mk_j)\|_\theta \xrightarrow{k\to 0} 0 \ .$$

Note that in finite dimensional spaces all norms are equivalent. Thus, for every fixed $j$ there exists a time step size $k(h_j)$, such that for all $k \le k(h_j)$ we can guarantee

$$e(h_j,k) \le \frac{1}{j} \ . \tag{5.68}$$

The achieved convergence in (5.68) holds for the modified problem (5.51) and its discretisation (5.53), where we inserted a cut-off function $\zeta_K$. We already showed in the proof of Theorem 5.2.8 that the semi-discrete problem is not altered by this modification. Now we address the modification of the fully discretised problem.

Remember that there is a constant $K$, such that the solution of the ODE (5.53), namely $U_{h_j}$, respects the barrier $K$ in (5.67). Therefore choosing the new cut-off barrier as $(K+1)$, it must also be possible to satisfy for $k(h_j)$ small enough that

$$\sup_{m\in\mathbb{N},0\le mk\le T_0} (\|T_{h_j}U^m_{h_j,k}\|_{\mathrm{C}} + |\mathbb{L}(T_{h_j}U^m_{h_j,k})|) \le K+1 \ .$$

This then justifies to undo our modification in (5.59) and to still obtain for all $k < k(h_j)$ the bound in (5.68) for the unmodified version. Together with the already achieved convergence in (5.37) and the simple inequality

$$\|T_{h_j}U^m_{h_j,k} - u(mk)\|_\theta \le \|T_{h_j}U^m_{h_j,k} - T_{h_j}U_j(mk)\|_\theta + \|T_{h_j}U_j(mk) - u(mk)\|_\theta \ , \tag{5.69}$$

this proves this corollary. □

## 5.3 Experiments and Discussion of Parameter Dependence

In this section we want to highlight the behaviour of solutions and the tendency to develop hysteretic behaviour. The properties shown in this section are dynamic effects and rely on the concurrent influence of different terms in the PDE (3.9).

Before we discuss the behaviour of solutions to (3.9) in numerical experiments, let us take a look at the PDE of interest again, but this time in non-divergence form

$$\partial_t u(t,x) + \partial_x \Big[ (p(t) + \frac{\mathbb{L}(u(t)) - \psi'(x)}{\tau}) u \Big] = \frac{\nu^2}{\tau} \partial_{xx} u(t,x) \ . \tag{5.70}$$

For the declaration of the non-flux boundary condition and the definition of $\mathbb{L}$ we refer to equations (3.9) and (3.8) in Chapter 3. For all following numerical experiments we use

$$\psi'(x) = 2x^3 - 3x^2 + \frac{6 \cdot 3 \cdot 7}{100} x \ . \tag{5.71}$$

This function is also plotted in Figure 5.1. In the same way as the chemical potential in Section 2.3, the potential $\psi'$ has two monotone increasing parts enclosing a decreasing part. In later discussions we will see, that this is crucial for the observed phenomena.

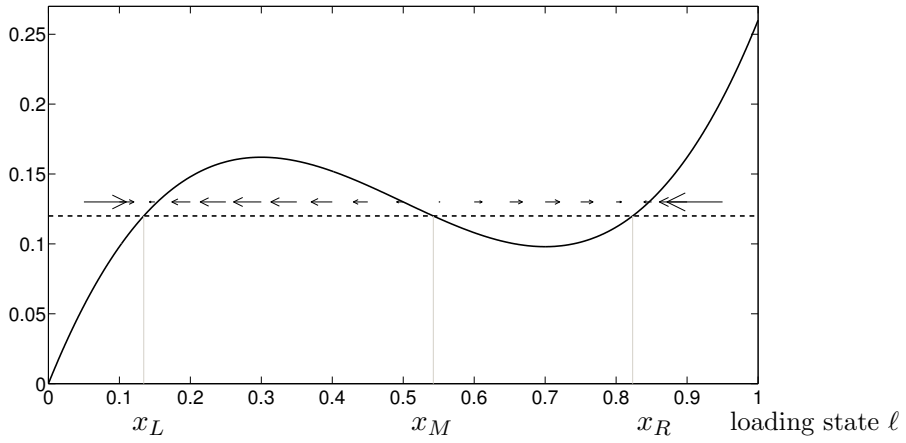



Figure 5.1: Velocity field of drift (arrows), induced by the difference of potential $\psi'$ (solid) and given value $\mathbb{L}(u(t))$ (example value of $\mathbb{L}(u(t)) = .12$ as dashed line).

The observed behaviour depends on the concurrent influence of the following three terms in the PDE and how they are weighted, depending on the choice of $\nu, \tau > 0$.

1. $\partial_x \Big[ p(t) u \Big]$ - This term induces a drift of mass which is equally strong in all $\Omega = ]0,1[$. It ensures that the first moment $\mathcal{C}(u(t))$, see (3.2), behaves as prescribed and $\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{C}(u(t)) = p(t)$ is satisfied.

2. $\partial_x \Big[ \frac{\mathbb{L}(u(t)) - \psi'(x)}{\tau} u \Big]$ - For given values of $\mathbb{L}(u(t))$, this term induces a drift of mass to certain points $x_0$ where $\psi'(x_0) = \mathbb{L}(u(t))$. In order for $x_0$ to be a point of attraction for this drift, $\psi'$ must be increasing at $x_0$ or in other words $\psi''(x_0) > 0$. Otherwise it would be a repelling point. A typical velocity field is depicted in Figure 5.1.

3. $\frac{\nu^2}{\tau} \partial_{xx} u(t,x)$ - This term induces diffusion, which would distribute the mass over

the whole domain $\Omega$ if the other terms were absent. It is the counterpart to the above drift, which wants the mass to concentrate at certain points.

(*Observation 1.*) The first notable observation to be mentioned is, that $\mathbb{L}(u(t))$ is dominated by the expected value of the chemical driving force $\psi'$

$$\mathbb{L}(u(t)) \approx \int_\Omega \psi'(x)u(t,x)\,\mathrm{d}x \qquad \Leftrightarrow \qquad \nu^2(u(1,t)-u(0,t)) \approx 0\ ; \tag{5.72}$$

compare to the definition of $\mathbb{L}$ in (3.8). This can be explained using two arguments. First, we are interested in solutions for small $\nu$. This gives a small pre-factor in (5.72). On the other hand we consider solutions for small $\tau$ and we prescribe the centre of mass $\mathcal{C}(u(t))$ to be away from the boundaries $x=0$ and $x=1$. This induces a drift away from the boundary, as described in the second term (see the enumeration of terms above). Due to this drift, the absolute value of the solution at the boundary gets small on a fast time scale, compare Figure 5.2. The experiments documented in this figure are conducted with $p \equiv 0$, but are typical for the general behaviour of solutions.

(*Observation 2.*) The solution tends to concentrate mass on a fast time-scale around two points $x_L < x_R$. For known value $\mathbb{L}(u(t))$ these two points are solutions to

$$\psi'(x_L(t)) = \mathbb{L}(u(t))\ , \qquad \psi''(x_L(t)) > 0\ , \tag{5.73}$$

$$\psi'(x_R(t)) = \mathbb{L}(u(t))\ , \qquad \psi''(x_R(t)) > 0\ . \tag{5.74}$$

Here comes the crucial form of $\psi'$ into play. Its form with two local extrema permits the existence of two solutions $x_L$ and $x_R$ in (5.73) and (5.74). Note that depending on the value of $\mathbb{L}(u(t))$ it is also possible that only one solution $x$ to $\psi'(x(t)) = \mathbb{L}(u(t))$ exists. In this case there is only one point of mass concentration. Two typical evolutions are shown in Figure 5.3 and Figure 5.4.

The width of these patterns mainly depends on the parameter $\nu$, as this is the weight of the third term in comparison to the second term. This can be seen in experiments documented in Figure 5.5. The initial value for this sequence of experiments was

$$u_0(x) = \begin{cases} c_0(x-.02)^2(x-.22)^2\ , & \text{for } x \in ]0.02, 0.22[\ . \\ 0\ , & \text{for } x \notin ]0.02, 0.22[ \cup ]0.72, 0.92[\ . \\ c_0(x-.72)^2(x-.92)^2\ , & \text{for } x \in ]0.72, 0.92[\ . \end{cases}$$

The constant $c_0$ is chosen such that $\int_\Omega u_0(x)\,\mathrm{d}x = 1$. The time scale on which this mass concentration takes place is mainly influenced by the choice of $\tau$. In order to measure this effect, we define for given value $\mathbb{L}(u(t))$ the point $x_M(t) \in ]x_L(t), x_R(t)[$ as the spatial position where $\psi'$ is decreasing and crossing the height $\mathbb{L}(u(t))$,

$$\psi'(x_M(t)) = \mathbb{L}(u(t))\ , \qquad \psi''(x_M(t)) < 0\ . \tag{5.75}$$

The two variances $\mathrm{var}_L$ and $\mathrm{var}_R$ defined in the following are then measurements of the

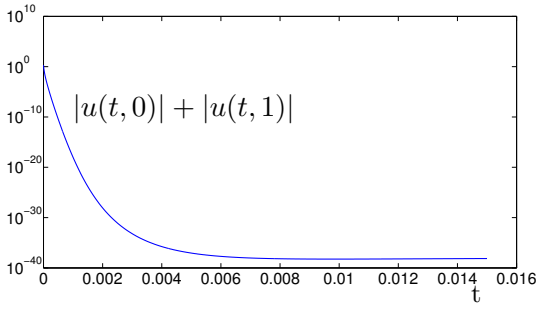



Figure 5.2: Typical evolution of absolute value of solutions at the boundaries, experimental data: $u_0 \equiv 1$, $p \equiv 0$, $\nu^2 = 10^{-4}$, $\tau = 10^{-3}$.

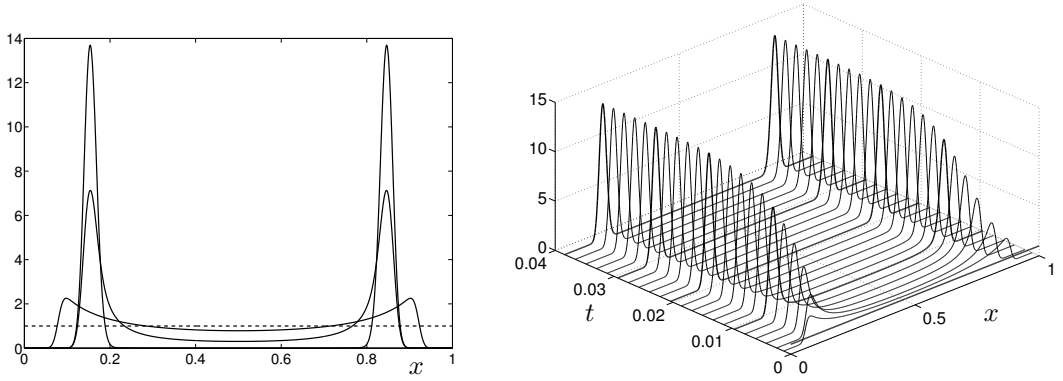



Figure 5.3: Typical evolution showing the concentration of mass around two points, left: solutions are plotted for times $t = 0.001, 0.005, 0.025$, right: visualisation of evolution of solutions , experimental data: initial data (dashed) is $u_0 \equiv 1$ , $p \equiv 0$, $\nu^2 = 10^{-4}$, $\tau = 10^{-3}$.

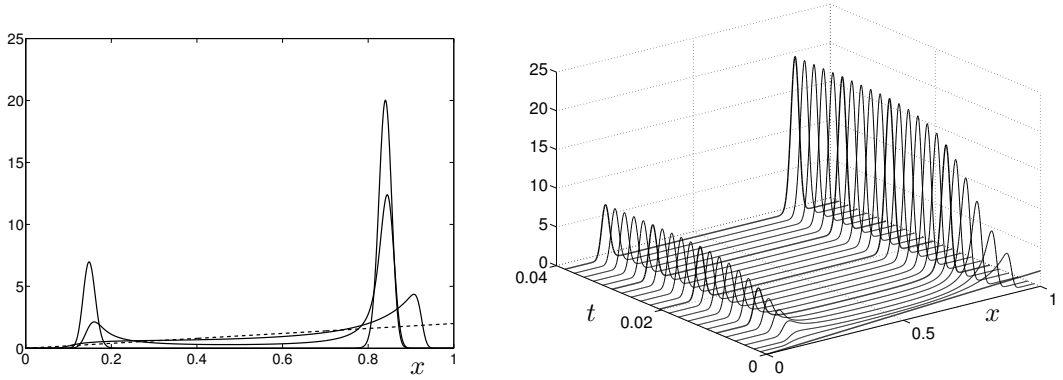



Figure 5.4: Typical evolution showing the concentration of mass around two points, left: solutions are plotted for times $t = 0.001, 0.005, 0.025$, right: visualisation of evolution of solutions , experimental data: initial data (dashed) is $u_0(x) = 2x$ on right, $p \equiv 0$, $\nu^2 = 10^{-4}$, $\tau = 10^{-3}$.

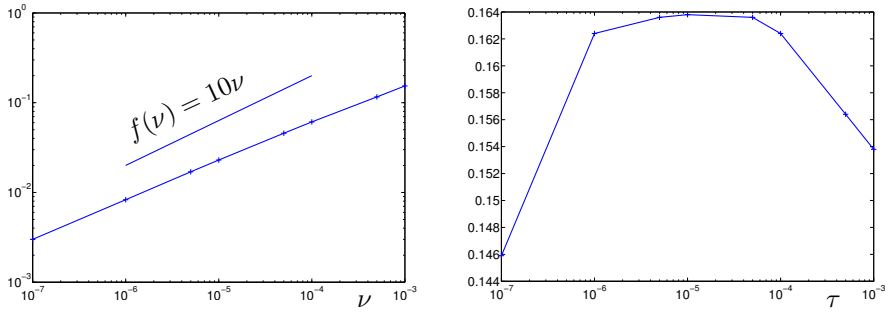



Figure 5.5: Width of pattern formed around a point, experimental data: initial data see text, left side: $\tau^2 = 10^{-3}$ and varying $\nu$, right side: $\nu = 10^{-3}$ and varying $\tau$, width is measured at time $t = 0.05$ and at height one on the left side.

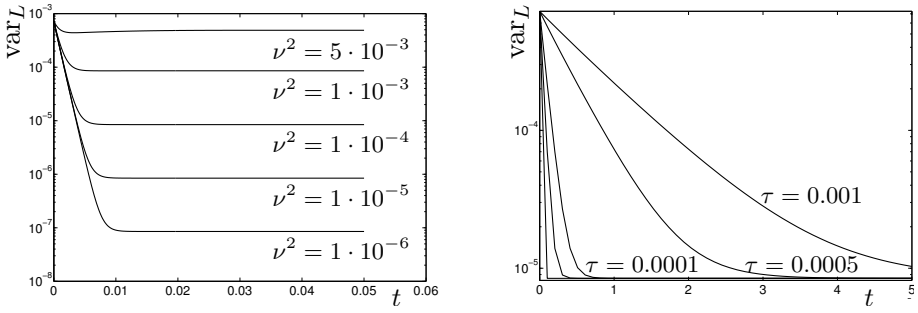



Figure 5.6: Speed of the mass concentration visualised in the evolution of $\mathrm{var}_L$ ($\mathrm{var}_R$ is similar), experimental data: initial data as in Figure 5.5, left side: $\tau^2 = 10^{-3}$ and varying $\nu$, right side: $\nu = 10^{-5}$ and $\tau = 10^{-3}, 5{\cdot}10^{-4}, 10^{-4}, 5{\cdot}10^{-5}, 10^{-5}$.

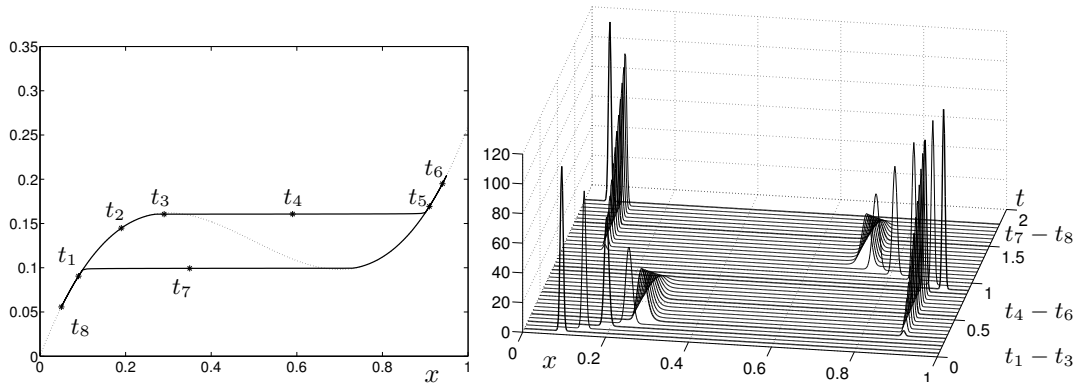



Figure 5.7: Typical hysteretic evolution behaviour of solutions, left: $\ell(t)$ vs $\langle\psi'\rangle(t)$ solid and $\psi'(x)$ dashed, right: evolutionary behaviour of solution, experimental data: initial data see solution at $t_1$, $\nu^2 = \tau = 10^{-5}$ and $p(t) = 1$ for $t_1$-$t_6$ and $p(t) = -1$ for $t_6$-$t_8$.

clustering of mass around $x_L$ and $x_R$,

$$\mathrm{var}_L(t) := \int_{(0,x_M)} (x - x_L)^2 u(t)\,\mathrm{d}x\ , \qquad \mathrm{var}_R(t) := \int_{(x_M,1)} (x - x_R)^2 u(t)\,\mathrm{d}x\ . \tag{5.76}$$

Their evolution in time, depicted in Figure 5.6, shows the dependency on $\tau$. The experiments documented in Figure 5.3 - 5.6 are made with data $p \equiv 0$. However, according to our experiments, they are typical for the general behaviour of solutions, when $p$ is of order one.

(*Observation 3.*) When the three terms (see page 75) of the PDE (5.70) are in balance, one can observe hysteretic behaviour of solutions. A typical occurrence of this hysteresis is shown in Figure 5.7. Therein we see a graph of the mean chemical potential,

$$\langle \psi' \rangle(t) := \int_\Omega \psi'(x) u(t,x)\,\mathrm{d}x\ ,$$

against the first moment $\mathcal{C}(u(t)) := \int_\Omega x u(t,x)\,\mathrm{d}x = \ell(t)$. The traversed path when $\mathcal{C}(u(t))$ is increasing differs to the path for decreasing $\mathcal{C}(u(t))$. This is also described in experiments by Dreyer, Guhlke and Herrmann in [11].

Let us discuss the fine balance of the parameters $\tau, \nu > 0$ and $p(t)$, which is needed in order to observe this hysteretic behaviour. We choose in our experiments $p(t) = 1$ or $p(t) = -1$ and vary the parameters $\tau$ and $\nu$. We do so since one can for any constant function $p(t) = c$ and a pair $\tau, \nu$ do a simple time transformation and get a resulting problem with some parameters $\tilde{\tau}, \nu$ and with $|\tilde{p}(t)| = 1$. This shows a first connection of the parameter and the *loading speed* $p$.

If we are in the regime $\nu^2 \ll \tau$, then we can see no hysteresis. The fast mass transport from left to right, as seen at the position 3 to 5 in Figure 5.7, vanishes. Instead, we see that the homogeneous drift from the first term, see page 75, dominates and the solution seems to make a simple translation to the right in Figure 5.8. It is astonishing, that a concentration of mass can persist around a point which is in the inner region $]x_*, x^*[$. This is similar to the delayed bifurcation effect mentioned in Section 2.3 and in [30].

As we described before, the second term induces a drift which widens such patterns in the region $]x_*, x^*[$. The crucial point is, that the (drift)velocity field is continuous. Furthermore it vanishes at a point $x_M \in ]x_*, x^*[$, see Figure 5.1. Thus, it is comparable small in a neighbourhood of of this unstable point. Lets try to explain the persistence of a profile in the critical region, see Figure 5.8. In agreement with *Observation 2* we see in a first phase a gathering around a point $x_L(t)$ within the interval $]0, x_*[$. The thickness of the area where most of the mass is gathered, depends on $\nu^2$. If this thickness is small enough, then the time needed to spread this mass concentration when being in the inner interval $]x_*, x^*[$ (in our case $x_* = .3$, $x^* = .7$, compare Figure 2.5(b) and Figure 5.1) is longer than the time to travel through it, while being pushed through by the first term $\partial_x\Big[p(t)u\Big]$. So we see it persisting. Remember that the thickness decreases when $\nu$ decreases and the drift producing the afore mentioned instability in $]x_*, x^*[$ gets weaker when $\tau$ increases.

The opposite can be observed in the parameter regime $\tau \ll \nu^2$. The diffusional term

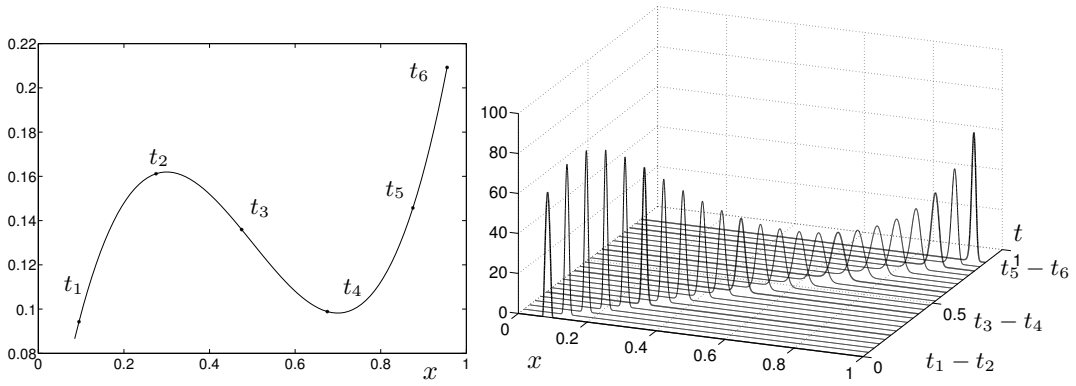



Figure 5.8: Delayed bifurcation effect for $\nu^2 \ll \tau$, left: evolution of $\langle \psi' \rangle$, right: evolution of solution, experimental data: initial data see solution at $t_1$, $\nu^2 = 1e^{-5}, \tau = 10^{-1}$ and $p(t) = 1$.

in (5.70) outweighs the other terms, and so the strongest mechanism in the PDE is the speed to equilibrium. Indeed assuming $p \equiv 0$ in (5.70) shows, that taking a sequence $\tau_k \to 0$ for constant $\nu$ can be interpreted as a time transformation from $t$ to $s = t/\tau_k$. This timetransformation then lets the factor $\tau_k$ disappear in (5.70). The time $t = 1$ then belongs to $s = 1/\tau_k \to \infty$. For bounded and nonzero $p$ we can argument formally that $p = \partial_t \ell$ transforms into $\tilde{p} = \partial_s \ell = \tau p \to 0$ and we can assume a similar behaviour of solutions to the case $p \equiv 0$. Thus the long-time behaviour of the PDE dominates. This is, according to Section 4.3, a convergence to stationary states. For details on the stationary solutions see Section 3.3. An experiment showing this behaviour is depicted in Figure 5.9. The solutions tends on a fast time scale to the equilibrium solution for $\ell(t_0)$. While $\ell(t) = \mathcal{C}(u(t))$ changes in time by $\dot{\ell}(t) = p(t)$, the solution follows approximately the steady states defined by the condition $\ell(t) = \mathcal{C}(u(t))$.

In [11], Dreyer et al. made a large number of numerical experiments and found evidence that the balance of parameters needed for the occurrence of hysteresis is $\tau \approx \nu^2$. For $\tau \to 0$ and $\nu \to 0$ in the regime $\tau \approx \nu^2$, one observes horizontal plateaus at the height of the local extrema, plotting the loading state $\mathcal{C}(u(t))$ versus the mean chemical driving force $\langle \psi' \rangle(t) := \int_\Omega \psi(x) u(t,x)\, \mathrm{d}x$, as shown in Figure 5.7. They also observed the other two limiting behaviours depicted in Figure 5.8 and Figure 5.9.

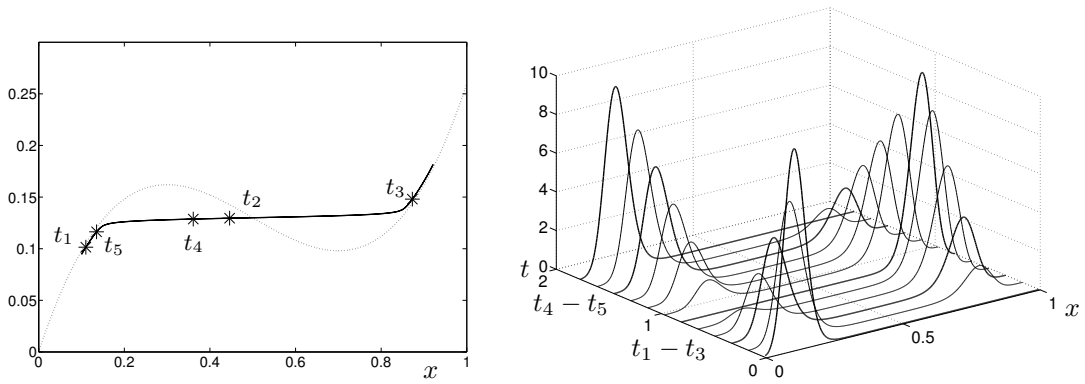



Figure 5.9: Effect of vanishing hysteresis for $\tau \ll \nu^2$, left side: $\ell(t)$ vs $\langle\psi'\rangle(t)$ solid and $\psi'(x)$ dashed, right: solutions at various times, corresponding to times $t_i$ as marked in left graph, experimental data: initial data see solution at $t_1$, $\nu^2 = 1e^{-3}, \tau = 10^{-7}$ and $p(t) = 1$ in first half of experiment and $p(t) = -1$ in the second half.

# 6 Formal Asymptotic Analysis for the Nonlinear Fokker–Planck Equation

We will now further investigate the influence of the parameters $\tau$ and $\nu$ in the PDE (3.9) to their solutions. This is of interest, because for the underlying model, i.e. the lithium ion battery, those parameters are relatively small. We will deduce further evidence for the tendency of solutions to concentrate mass around certain points with the method of *matched asymptotic expansions* and derive approximations for the solution. This is in accordance with numerical experiments in Section 5.3, where we observed this behaviour as well.

In order to clarify the meaning of *relatively small* parameters we inspect the dimensions of the quantities in the Fokker–Planck equation (3.9). We repeat this PDE here, written in dimensionless variables,

$$\frac{\tau}{\nu^2}\partial_t u(t,x) = \partial_{xx} u(t,x) + \partial_x\Big(u(x)\big[\frac{\Psi}{\nu^2}\psi'(x) - \mathbb{L}(u(t)) - \frac{\tau}{\nu^2}\dot{\ell}(t)\big]\Big)\,, \qquad x \in ]0,1[\,,$$
$$\mathbb{L}(v)) = (v(1) - v(0)) + \int_\Omega \frac{\Psi}{\nu^2}\psi'(x)v(x)\,\mathrm{d}x\,, \qquad \text{for all } v \in \mathrm{H}^1(\Omega)\,,$$

with the constants

| | |
|---|---|
| $\tau$ | time scale, such that $\dot{\ell}$ is of order 1 , |
| $\nu^2$ | material constant, |
| $\Psi$ | scale for the potential $\psi$, such that $\psi'$ is of order 1 . |

The demand of $\psi'$ being of order one refers to the values in an interval of interest. We mention this because $\psi'$ may go to infinity as it approaches the boundary of $[0,1]$. Our choice for this interval of interest is $[x_{**}, x^{**}]$, see Figure 2.5(b), and so according to [11] we find

$$\Psi = 10^{-1}, \qquad \nu^2 \in [10^{-8}, 10^{-5}], \qquad \tau \in [10^{-7}, 10^{-3}]. \tag{6.1}$$

and hence, $\tau/\nu^2 \in [10^{-2}, 10^{5}]$ , $\Psi/\nu^2 \in [10^{-7}, 10^{-4}]$. Note that we do not introduce a scaling for $u$, as this shall represent a probability density, i.e. a positive function with mean value one, and thus does not permit rescaling.

The method of *matched asymptotic expansions* is dedicated to derive approximations for singular perturbed problems, for example to derive jump conditions at phase boundaries as the limit of phase field models where a certain parameter tends to zero. Roughly spoken, one assumes for a family of solution $g_\nu$ to a problem with a parameter $\nu$ that $g_\nu$

can be decomposed in two ways. First, one separates the solution spatially by considering subdomains $\omega_k$, such that their union covers the whole considered domain $\Omega$. On each subdomain $\omega_k$ one then assumes the possibility to further decompose the solution in one of two ways. The first is called the *outer expansion* where the solution is written as a sum in the form

$$g(\nu, t, x) = R_\alpha(\nu, t, x) + \sum_j \xi_{\alpha,j}(\nu) g_{\alpha,j}(t, x) \ , \qquad \text{for } x \in \omega_\alpha \ .$$

The functions $g_{\alpha,j}$ do not depend on $\nu$. In the second variant, the *inner expansion*, a further spatial transformation $\phi$ comes into play,

$$g(\nu, t, x) = R_\alpha(\nu, t, x) + \sum_j \xi_{\alpha,j}(\nu) G_{\alpha,j}(t, \phi(\nu, x)) \ , \qquad \text{for } x \in \omega_\alpha \ .$$

One assumes that the remainder $R_\alpha$ is small or ideally zero. In both cases, partial sums serve as approximations to the original solution. Depending on the underlying problem one derives properties for the functions $g_{\alpha,j}$ and $G_{\alpha,j}$. In general it is not known a priori, whether such a representation as a sum exists, or what the coefficients $\xi$ look like. Nevertheless, this method is accepted in order to find evidence for limiting behaviour, when rigorous methods are not fruitful. The outcome usually serves as a (formal) justification of the choice made for the coefficients. For this reason we have to point out explicitly that the calculations shown in this section have a formal character. Thus, the results can only serve as an indication or evidence but not as rigorous results. In the following we have to rely on assumptions which are yet, to the point of writing, open and we will point out those assumptions explicitly.

Motivated by numerical experiments in Section 5.3, we want to investigate the limiting behaviour for solutions to the PDE (3.9), for the case of choosing the parameters

$$\beta\tau = \nu^2 \ ,$$

and considering sequences $\nu_k \to 0$ for fixed $\beta > 0$. First, we will reformulate the PDE

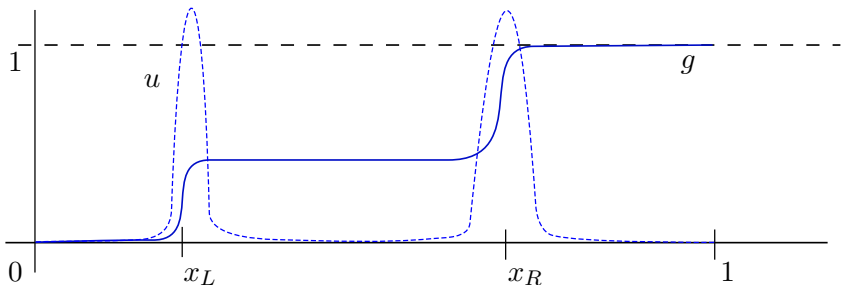



Figure 6.1: The function $g$ (solid) develops profiles similar to jumps at positions $x_L$ and $x_R$, where $u = \partial_x g$ (dashed) concentrates mass.

(3.9) by passing to a new variable $g$, defined as

$$g(\nu, t, x) := \int_0^x u(\nu, t, y)\,\mathrm{d}y \ ,$$

where $u$ is the solution to (3.9), when choosing the parameters $\tau$ and $\nu$ as above. We explicitly account for the dependence on $\nu$ by naming it in the list of arguments. By this transformation, we get the equivalent PDE

$$\begin{aligned} \partial_t g(\nu,t,x) + \frac{\beta}{\nu^2}\Big[\mathbb{L}(\partial_x g(\nu,t)) + \frac{\nu^2}{\beta}\dot{\ell}(t) - \psi'(x)\Big]\partial_x g(\nu,t,x) &= \beta\partial_{xx} g(\nu,t,x) \ , \\ g(\nu,0,t) = 0 \ , \qquad g(\nu,1,t) &= 1 \ , \\ \mathbb{L}(v) = \nu^2(v(1) - v(0)) + \int_\Omega \psi'(x)v(x)\,\mathrm{d}x \ , & \end{aligned} \tag{6.2}$$

combined with transformed initial values. Note that we already inserted the assumed dependence of $\tau$ on $\nu$. The function $p = \dot{\ell}$ shall be a piecewise Hölder continuous function of time and $\psi \in \mathrm{H}^1(\Omega)$ is a double well as described in Section 2.3. If $u(t)$ tends to concentrate mass at a certain position, then the function $g(t)$ will tend to develop a profile similar to a jump, as depicted in Figure 6.1.

We will now derive approximations for the solution to (6.2) in five subdomains, see Figure 6.2. Two domains, $\omega_L$ and $\omega_R$, are neighbourhoods of the two critical points $x_L(t)$ and $x_R(t)$, where we expect a jumping behaviour. The other domains $\omega_1$, $\omega_2$ and $\omega_3$ are the remaining parts of $\Omega$ in between (with an additional overlap). The domains may also change in time, as the positions $x_L(t)$ and $x_R(t)$ do. The precise definition of the subdomains depends on the parameter $\nu$, as we shall see in Section 6.3.

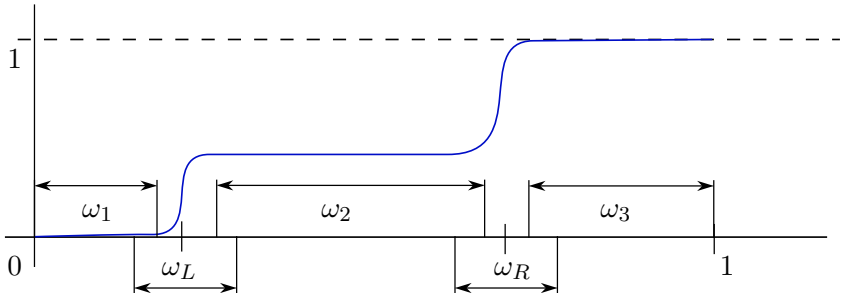



Figure 6.2: Sketch of the layout of the different domains of validity where the solution will be approximated by differing expansions in each subdomain.

## 6.1 Inner Expansion

In this section we inspect the solution $g$ in the domains $\omega_L$ and $\omega_R$. As the treatment in both cases is identical, we add the index $\alpha$ to variables and constants. This shall be

read as either $\alpha = L$ or $\alpha = R$ and indicates quantities that may vary in both cases. We want to account for the possibility that the solution may show a large variation in a small neighbourhood of a yet unknown position $x_\alpha$. We expect those positions to be the centre of the pulses of $u$, see Figure 6.1. Later we will determine their positions depending on the value of $\int_\Omega \psi'(x)u(x,t)\,\mathrm{d}x$. In order to account for a transition of the function values of $g$ in a small interval, we transform the spatial variable $x$ into $z$ in the form

$$z = \frac{x - x_\alpha}{\nu}, \qquad \alpha \in \{L, R\} \ . \tag{6.3}$$

Then around the critical points $x_\alpha(t)$ we define $G_\alpha$ as

$$\begin{aligned} g(\nu,t,x) = g(\nu,t,x_\alpha + \nu z) =: G_\alpha(t,z) \ , \\ \Rightarrow \partial_t g(\nu,t,x) = \partial_t G_\alpha(t,x) - \frac{\dot{x_\alpha}}{\nu}\partial_z G_\alpha(\nu,t,z), \qquad \partial_x(\cdots) = \nu^{-1}\partial_z(\cdots) \ . \end{aligned}$$

Now we make the formal assumption that we can write $G_\alpha$ as a power series in the form

$$G_\alpha(\nu,t,z) = \sum_{k=0}^{\infty} \nu^k G_{\alpha,k}(t,z) \ . \tag{6.4}$$

Furthermore, $\partial_z G_\alpha$ shall be the same series when replacing $G_{\alpha,k}$ by $\partial_z G_{\alpha,k}$. Note that the functions $G_{\alpha,k}$ do not depend on $\nu$.

We introduce $\Lambda(t,\nu) = \mathbb{L}(\partial_x g(t,\nu)) + \nu^2/\beta\ \dot{\ell}(t)$ and assume that it is possible to write $\Lambda(t,\nu)$ as a series in $\nu$ as well,

$$\Lambda(t,\nu) := \mathbb{L}(\partial_x g(t,\nu)) + \frac{\nu^2}{\beta}\dot{\ell}(t) = \sum_{k=0}^{\infty} \nu^k \Lambda_k(t) \ . \tag{6.5}$$

Additionally, we assume that $\psi'$ can be written as a Taylor series in the neighbourhood around $x_\alpha$,

$$\psi'(x) = \psi'(x_\alpha + \nu z) = \sum_{k=0}^{\infty} \nu^k z^k \frac{1}{k!}\psi^{(k+1)}(x_\alpha) \ . \tag{6.6}$$

We insert the newly introduced quantities in (6.2) and get

$$0 = \partial_t G_\alpha - \nu^{-1}\dot{x_\alpha}\partial_z G_\alpha + \nu^{-3}\beta(\Lambda - \psi'(x_\alpha + \nu z))\partial_z G_\alpha - \nu^{-2}\beta\partial_{zz}G_\alpha \ .$$

where we dropped the dependencies on the arguments $z$, $t$ and $\nu$ in $G_\alpha$, $\Lambda$ and $x_\alpha$ for better readability. The crucial point is that this equality should hold for all $\nu$. Hence, we insert the Taylor series in (6.4), (6.5) and (6.6) and sort the terms according to their powers of $\nu$. Equating the corresponding coefficients, we get in the lowest two orders

the resulting ODE's:

$$\begin{aligned}
&[\text{order } \nu^{-3}] \qquad 0 = (\Lambda_0 - \psi'(x_\alpha))\partial_z G_{\alpha,0} \ , \\
&[\text{order } \nu^{-2}] \qquad 0 = (\Lambda_0 - \psi'(x_\alpha))\partial_z G_{\alpha,1} + (\Lambda_1(t) - z\psi''(x_\alpha))\partial_z G_{\alpha,0} - \partial_{zz} G_{\alpha,0} \ .
\end{aligned}$$

Finally, we determine $x_\alpha$, $\alpha \in \{L, R\}$, for given $\Lambda_0(t)$ as the two solutions to

$$\psi'(x_\alpha) = \Lambda_0(t) \qquad \text{and} \qquad \psi''(x_\alpha) > 0 \ . \tag{6.7}$$

This guarantees, that we do not get trivial solutions for $G_{\alpha,0}$ in the above ODEs. It should be mentioned, that in the case of the existence of only one solution, one is simply restricted to only one choice for $\alpha \in \{R, L\}$. By our definition of $x_\alpha(t)$ we get an ODE for $G_{\alpha,0}$ resulting from the terms of order $\nu^{-2}$,

$$0 = (\Lambda_1(t) - z\psi''(x_\alpha))\partial_z G_{\alpha,0} - \partial_{zz} G_{\alpha,0} \ .$$

In order to write the solution of this ODE, we introduce the yet unknown parameter $C_{\alpha,0}$ and $\lambda_\alpha$ and define for $\psi''(x_\alpha(t)) > 0$ the function

$$U_{\alpha,t}(z) := \sqrt{\frac{\psi''(x_\alpha(t))}{2\pi}} \exp\Big( - \frac{\psi''(x_\alpha(t))}{2} \ (z - \frac{\Lambda_1(t)}{\psi''(x_\alpha(t))})^2 \Big) \ .$$

Note that $\int_{-\infty}^{+\infty} U_\alpha(z)\,\mathrm{d}z = 1$. From the above ODE and the choice of $x_\alpha(t)$ we deduce

$$G_{\alpha,0}(z) = C_{\alpha,0} + \lambda_\alpha \int_{-\infty}^{z} U_{\alpha,t}(s)\,\mathrm{d}s \ . \tag{6.8}$$

The constants $C_{\alpha,0}$ and $\lambda_\alpha$ may also depend on time. In Section 6.3 we will determine the unknown constants appearing here by the method of matching.

## 6.2 Outer Expansion

The three remaining subdomains $\omega_1$, $\omega_2$ and $\omega_3$ are the areas away from the two critical points $x_L$ and $x_R$. We use the index $\alpha$ to represent either $\alpha = 1$, $\alpha = 2$ or $\alpha = 3$, depending on which subdomain $\omega_\alpha$ we consider $g$. This allows us to treat all three cases at once. In each subdomain respectively, we assume that solution $g$ to (6.2) can be extended as a series in $\nu$ in the form

$$g(\nu, t, x) = \sum_{k=0}^{\infty} \nu^k g_{\alpha,k}(t, x) \ . \tag{6.9}$$

As in (6.5) we assume that we can write $\Lambda(t)$ as a power series and insert this together with (6.9) into (6.2) to get

$$0 = \sum_{k=0}^{\infty} \nu^k \partial_t g_{\alpha,k}(t,x) + \beta\nu^{-2}(\sum_k \Lambda_k(t) - \psi'(x)) \sum_{k=0}^{\infty} \nu^k \partial_x g_{\alpha,k}(t,x) - \beta \sum_{k=0}^{\infty} \nu^k \partial_{xx} g_{\alpha,k}(t,x) \ .$$

We demand that this is fulfilled for all $\nu$. Thus, we sort the terms by powers of $\nu$. Equating the corresponding coefficients we deduce in the first order

$$0 = (\Lambda_0 - \psi'(x))\partial_x g_{\alpha,0} \ , \qquad \Longrightarrow g_{\alpha,0} \equiv c_{\alpha,0} \ ,$$

for yet unknown constants $c_{\alpha,0}$. These constant may depend on time, and are determined in the next section.

## 6.3 Matching

In the process of *matching* one exploits that the partial sums on all subdomains are approximations of the same functions. This serves as the source of additional information to determine the unknown constants which arose in the calculations in the Sections 6.1 and 6.2. We will use the method of matching as described by Lagerstrom in [25, Sec. 1.4].

We assume that, on their respective domains the functions $G_{L,0}$ and $G_{R,0}$ resulting from the inner expansions are good approximations (when transformed back from $z$ to $x$, see (6.3)) of the actual solution. The domain $\omega_L$ and $\omega_R$ where this is a valid approximation, may depend on the parameter $\nu$. We call this the *domains of validity*. Similarly we assume for $g_{1,0}$, $g_{2,0}$ and $g_{3,0}$ resulting from the outer expansions, that they have domains of validity $\omega_k$, $k = 1, 2, 3$ which may depend on $\nu$ as well. According to Lagerstrom we say that these domains must overlap and here we use the procedure of matching.

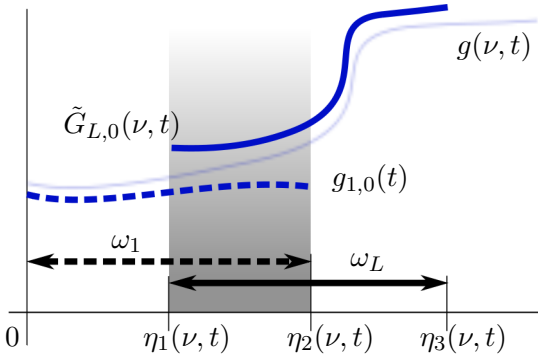



Figure 6.3: Sketch of the overlap of two domains of validity.

Let us explain the method of matching using the example of $G_{L,0}$ and $g_{1,0}$. We define the function $\tilde{G}_{L,0}$ by reversing the transformation of the variable in the inner expansion, see (6.3), and thus

$$\tilde{G}_{L,0}(\nu,t,x) = G_{L,0}(\frac{x-x_L(t)}{\nu},t) \ .$$

We refer to the Figure 6.3 for a sketch of the domains of validity $\omega_1 = ]0,\eta_2(\nu,t)[$ and $\omega_1 = ]\eta_1(\nu,t),\eta_2(\nu,t)[$. Their shape depends on $\nu$ and $t$. They overlap on

$$\omega_1 \cap \omega_L := ]\eta_1(\nu,t),\eta_2(\nu,t)[ \ .$$

On the *domain of validity* and the considered time interval $S \subset \mathbb{R}_+$, there shall hold

$$\lim_{\nu\to 0} \sup_{x\in\omega_1(\nu,t),\ t\in S} |g(\nu,t,x) - g_{1,0}(t,x)| \ = \ \lim_{\nu\to 0} \sup_{x\in\omega_L(\nu,t),\ t\in S} |g(\nu,t,x) - \tilde{G}_{L,0}(t,x)| = 0 \ ,$$

where $g$ is the exact solution to (6.2). Hence, this gives in the overlap area

$$\lim_{\nu\to 0} \sup_{x\in]\eta_1(\nu,t),\eta_2(\nu,t)[,t\in S} |\tilde{G}_{L,0}(t,x) - g_{1,0}(t,x)| = 0 \ ,$$

or simply for any function $\eta(\nu,t) \in ]\eta_1(\nu,t),\eta_2(\nu,t)[$ there should hold

$$\lim_{\nu\to 0} \sup_{t\in S} |\tilde{G}_{L,0}(t,\eta(\nu,t)) - g_{1,0}(t,\eta(\nu,t))| = 0 \ . \tag{6.10}$$

This relates two neighbouring approximations of the solution with each other. Deducing information from this is called matching. This is a demonstration for the matching in first order. As a remark we note, that in the next order one would ask for

$$\lim_{\nu\to 0} \sup_{t\in S} \frac{|\tilde{G}_{L,0}(\nu,t,\eta) + \nu\tilde{G}_{L,1}(\nu,t,\eta) - \Big(g_{1,0}(\nu,t,\eta) + \nu g_{1,1}(\nu,t,\eta)\Big)|}{\nu} = 0 \ .$$

The domains of validity and especially their dependence on $\nu$, may differ depending what order we inspect and want to match. For more explanation of matching in general orders we refer to Lagerstrom, [25, Sec. 1.4].

Let us proceed from (6.10). We do not give exact formulas for the functions $\eta_k$, $k = 1,2,3$, but we demand that for all times $t \in S$ the following identities must hold

$$\begin{aligned} \lim_{\nu\to 0} \eta_1(\nu,t) = \lim_{\nu\to 0} \eta_2(\nu,t) = x_L(t) \ , \\ \lim_{\nu\to 0} \frac{\eta_1(\nu,t) - x_L(t)}{\nu} = \lim_{\nu\to 0} \frac{\eta_2(\nu,t) - x_L(t)}{\nu} = -\infty \ . \end{aligned}$$

Thus, the equation (6.10) results in

$$\lim_{x\to x_L} g_{1,0}(t,x) = \lim_{z\to-\infty} G_{L,0}(z,t) \ , \qquad \Longrightarrow \qquad c_{1,0}(t) = C_{L,0}(t) \ .$$

At the same time we demand that the left boundary condition $g(\nu, t, 0) = 0$ is satisfied at all times and all $\nu$, which gives

$$0 = g(\nu, t, 0) = \sum_{k=0}^{\infty} \nu^k g_{1,k}(t, 0) = c_{1,0}(t) + \sum_{k=1}^{\infty} \nu^k g_{1,k}(t, 0) \ .$$

Hence, we deduce $c_{1,0}(t) = 0$ for all times.

We repeat this method of matching for all neighbouring approximations. Let us recapitulate the form of the first order approximations and the occurring unknowns:

$$g_{\alpha,0}(t, x) = c_{\alpha,0}(t) \ , \quad \text{for } \alpha = 1, 2, 3 \text{ and} \tag{6.11}$$

$$G_{\alpha,0}(z, t) = C_{\alpha,0}(t) + \lambda_\alpha(t) \int_{-\infty}^{z} U_{\alpha,t}(s) \, \mathrm{d}s \ , \quad \text{for } \alpha = L, R \ . \tag{6.12}$$

We report here only the results, as the calculations are a repetition of the procedure above.

$$\begin{aligned}
\text{matching with left boundary :} \quad & c_{1,0}(t) \equiv 0 \ , \\
\text{matching } g_{1,0} \text{ and } G_{L,0} : \quad & C_{L,0}(t) \equiv c_{1,0}(t) \equiv 0 \ , \\
\text{matching } G_{L,0} \text{ and } g_{2,0} : \quad & c_{2,0}(t) \equiv \lambda_L(t) \ , \\
\text{matching } g_{M,0} \text{ and } G_{R,0} : \quad & C_{L,0}(t) \equiv c_{2,0}(t) \equiv \lambda_L(t) \ , \\
\text{matching } G_{R,0} \text{ and } g_{3,0} : \quad & c_{3,0}(t) \equiv \lambda_L(t) + \lambda_R(t) \ , \\
\text{matching with left boundary :} \quad & 1 \equiv c_{3,0}(t) \equiv \lambda_L(t) + \lambda_R(t) \ .
\end{aligned}$$

We write $\lambda(t) := \lambda_R(t)$ and $\lambda_L(t) = 1 - \lambda(t)$. Summarising the results we know for the functions $g_{\alpha,0}$ and $G_{\alpha,0}$,

$$g_{1,0}(t, x) = 0 \ , \qquad g_{2,0}(t, x) = 1 - \lambda(t) \ , \qquad g_{3,0}(t, x) = 1 \ ,$$

$$G_{L,0}(z, t) = (1 - \lambda(t)) \int_{-\infty}^{z} U_{L,t}(s) \, \mathrm{d}s \ , \qquad G_{R,0}(z, t) = (1 - \lambda(t)) + \lambda(t) \int_{-\infty}^{z} U_{R,t}(s) \, \mathrm{d}s \ .$$

In order to derive more information, we now want to use the knowledge that solutions to the original PDE (3.9) satisfy $\mathcal{C}(u(t)) = \ell(t)$ for a prescribed function $\ell$. For the PDE (6.2) and its solution $g$, this gives rise to

$$1 - \ell(t) = \int_\Omega g(t, x, \nu) \, \mathrm{d}x \ .$$

We split this into integrals over subdomains, as sketched in to Figure 6.2,

$$\begin{aligned} 1-\ell(t) = &\int_0^{x_L^-(\nu)} g(\nu,t,x)\,\mathrm{d}x + \int_{x_L^+(\nu)}^{x_R^-(\nu)} g(\nu,t,x)\,\mathrm{d}x + \int_{x_R^+(\nu)}^{1} g(\nu,t,x)\,\mathrm{d}x \\ &+ \int_{x_L^-(\nu)}^{x_L^+(\nu)} g(\nu,t,x)\,\mathrm{d}x + \int_{x_R^-(\nu)}^{x_R^+(\nu)} g(\nu,t,x)\,\mathrm{d}x \ . \end{aligned} \tag{6.13}$$

The boundaries $x_L^- < x_L < x_L^+$ and $x_R^- < x_R < x_R^+$ are chosen in a way, that they lie in the overlap of the suitable domains of validity. Remember that $g_{\alpha,0}$ and $\tilde{G}_{\alpha,0}$ represent the approximations of $g$ in the first order of $\nu$. Hence, we deduce

$$\begin{aligned} 1-\ell(t) = &\int_0^{x_L^-(\nu)} g_{1,0}(t,x)\,\mathrm{d}x + \int_{x_L^+(\nu)}^{x_R^-(\nu)} g_{2,0}(t,x)\,\mathrm{d}x + \int_{x_R^+(\nu)}^{1} g_{3,0}(t,x)\,\mathrm{d}x \\ &+ \int_{x_L^-(\nu)}^{x_L^+(\nu)} G_{L,0}\Big(\frac{x-x_L(\nu)}{\nu},t\Big)\,\mathrm{d}x + \int_{x_R^-(\nu)}^{x_R^+(\nu)} G_{R,0}\Big(\frac{x-x_R(\nu)}{\nu},t\Big)\,\mathrm{d}x \quad + \ \mathcal{O}(\nu) \ , \end{aligned}$$

which holds for all $\nu$ and especially for $\nu \to 0$. We employ

$$x_{L,R}^-(\nu) \xrightarrow{\nu\to 0} x_{L,R}(t) \qquad \text{and} \qquad x_{L,R}^+(\nu) \xrightarrow{\nu\to 0} x_{L,R}(t)$$

and the fact that $G_{\alpha,0}$ has a bounded absolute value to get

$$\begin{aligned} & 1-\ell(t) = 0\ \big(x_L(t)\big) \ + \ \big(1-\lambda(t)\big)\big(x_R(t)-x_L(t)\big) \ + \ 1\big(1-x_R(t)\big) \\ \Rightarrow \qquad & \ell(t) = \big(1-\lambda(t)\big)x_L(t) + \lambda(t)x_R(t) \ . \end{aligned}$$

This is the same as condition (2.11), which we deduced from the discrete many particle model in Section 2.3. Furthermore the two filling degrees are connected by (6.7), which is equivalent to (2.12). Hence, by Lemma 2.3.5 the solution to the Fokker-Planck equation (3.9) can be characterised asymptotically in the limit of $\nu^2 = \beta\tau \to 0$ by two filling degrees $x_L(t)$ and $x_R(t)$, together with the phase fraction $\lambda(t)$ describing the splitting of mass into two delta distributions.

Let us investigate the function $\Lambda(t,\nu)$, see (6.5), and its dependence on $\nu$. Using its definition (6.5) and the definition of $\mathbb{L}$ in (3.8) as well as $\beta\tau = \nu^2$, we can write

$$\Lambda(t,\nu) = \int_\Omega \psi'(x)\partial_x g(\nu,t,x)\,\mathrm{d}x + \nu^2\big(\partial_x g(\nu,t,1) - \partial_x g(\nu,t,0)\big) + \frac{\nu^2}{\beta}\dot{\ell}(t) \ .$$

Splitting the integrals as in the last calculations, see (6.13), we easily get

$$\Lambda_0(t) = \psi'(x_R(t)) = \psi'(x_L(t)) \ , \tag{6.14}$$

which is consistent with the definition of $x_\alpha$ in (6.7). What we want to point out is that the value $\Lambda$ (and $\mathbb{L}$) is dominated by the influence of the first term, i.e. the integral, whereas the other parts are of higher order with respect to $\nu$. This is in agreement to

observations in experiments in Section 5.3.

The investigation of the evolution of $\lambda$ brings us to higher order approximation in the domain $\omega_2$, see Figure 6.2. Continuing our calculation for the outer expansion gives in the next two orders

$$\begin{aligned} 0 = (\Lambda_0 - \psi'(x))\partial_x g_{\alpha,1} \ , \qquad & \Longrightarrow g_{\alpha,1} \equiv c_{\alpha,1} \ , \\ \partial_t g_{\alpha,0} = -\beta(\Lambda_0 - \psi'(x))\partial_x g_{\alpha,2} \ , \qquad & \Longrightarrow \partial_x g_{\alpha,2}(x,t) = \frac{-\partial_t g_{\alpha,0}}{\beta(\Lambda_0 - \psi'(x))} \ . \end{aligned}$$

We do not want to match these with the neighbouring expansions and we also do not care for the value of $c_{\alpha,1}$ which might depend on time. The interesting part here is $\partial_x g_{2,2}(x,t)$. Remember that $g_{2,0} \equiv 1 - \lambda(t)$ and hence

$$\partial_x g_{2,2}(x,t) = \frac{\frac{\mathrm{d}}{\mathrm{d}t}\lambda(t)}{\beta(\Lambda_0 - \psi'(x))} \ .$$

This function has singularities at the points $x_R(t)$ and $x_L(t)$, see (6.7) and (6.14). Therefore the domain of validity $\omega_2$ in which $g = g_{2,0} + \nu g_{2,1} + \nu^2 g_{2,2} + \mathcal{O}(\nu^2)$, must be smaller than $]x_L(t), x_R(t)[$.

As we explained before, the domain $\omega_2$ depends on $\nu$ (and time) and we assume that the boundaries of $\omega_2$ converge to $x_L(t)$ and $x_R(t)$ as $\nu \to 0$. We deduce from this that $\frac{\mathrm{d}}{\mathrm{d}t}\lambda(t) = 0$ if both inequalities, $x_L(t) \neq x_*$ and $x_R(t) \neq x^*$, are satisfied. Indeed assume, that $\frac{\mathrm{d}}{\mathrm{d}t}\lambda(t) \neq 0$ and $x_L(t) \neq x_*$, then by $\Lambda_0(t) = \psi'(x_L(t))$ we know that the function $(\Lambda_0 - \psi'(x))$ changes its sign in $]x_L, x_R[$, see Figure 2.5(b). Hence $(\Lambda_0 - \psi'(x))$ changes its sign also in $\omega_2$ if $\nu$ is small enough. That means that $\partial_x g_{2,2}(x,t)$ is negative and consequently $g_{2,2}$ is decreasing somewhere in $\omega_2$. We know that $g_{2,0}$ and $g_{2,1}$ are constant functions in $\omega_2$ and the slope of $g = g_{2,0} + \nu g_{2,1} + \nu^2 g_{2,2} + \mathcal{O}(\nu^2)$ is thus dominated by $g_{2,2}$. But then, for small $\nu$, $g$ must be decreasing somewhere in $\omega_2$ as well. This however is a contradiction to $u = \partial_x g$ being nonnegative as shown in Lemma 4.1.5. Thus $\frac{\mathrm{d}}{\mathrm{d}t}\lambda(t) = 0$ as long as $x_L < x_*$ and $x_R > x^*$.

## 6.4 Summary of Results of Matched Asymptotic Analysis

We saw, that the first order approximation of the formal asymptotic analysis predicts a mass concentrating behaviour around critical points $x_L(t)$ and $x_R(t)$. Furthermore, the position of these points can be determined by the phase fraction $\lambda(t)$ and the loading state $\ell(t)$, as described in Lemma 2.3.5. Hence, it is very similar to the limit problem in Section 2.3, where the system is described by the evolution of two phases.

The first order approximation of $g$ being the solution to (6.2), tends to have jumps at the points $x_\alpha$ where $\psi'(x_\alpha) = \int_\Omega \psi'(x)u(t,x)\,\mathrm{d}x$. At the same time $g$ is the primitive of $u$ which is the solution to the original problem (3.9). Hence, a function $g$ tending to jump means, that $u$ tends to delta distributions at these points. We can even say that for small $\nu$ the solution $u$ has a form similar to a Gaussian distribution concentrated

around $x_L(t)$ and $x_R(t)$,

$$u(x,t) \approx \lambda_\alpha(t)\sqrt{\frac{\psi''(x_\alpha(t))}{2\pi\nu^2}}\ \exp\Big(-\frac{\psi''(x_\alpha(t))}{\nu^2}(x-x_\alpha(t))^2\Big)\ ,$$

for either $\alpha = L$ or $\alpha = R$ and $\lambda_R(t) = \lambda(t)$, $\lambda_L(t) = 1 - \lambda(t)$. This can be seen by the given formula of $U_{\alpha,t}$ in (6.8) and assuming that $\Lambda_1(t) \equiv 0$. Away from $x_L$ and $x_R$ the function $u$ is approximately zero, because $g$ is approximately constant in the first order. The quality of this approximation can be observed in Figure 6.4 and Figure 6.5. We see a good approximation quality of the patterns at critical positions $x_\alpha$ away from the positions $x_* = .3$ or $x^* = .7$. Remember from Figure 2.5(b) or Figure 5.1, that at these positions $\psi' = 0$. In Figure 6.5 (top) we see, that close to $x_*$ the quality of the approximation drops. At this position mass transfer from left to right occurs and the behaviour can probably not be captured satisfactorily with a simple first order approximation any more.

Unfortunately, the evolution of the phase fraction $\lambda(t)$ for the limiting two phase system is still open. We saw in the end of the last section, that whenever the positions $x_L(t)$ and $x_R(t)$ satisfy $x_L < x_*$ and $x_L > x^*$ then $\frac{\mathrm{d}}{\mathrm{d}t}\lambda(t) = 0$. This is in agreement with the behaviour of the discrete many particle system in Section 2.3. However the case $x_L = x_*$ or $x_R = x^*$ are not covered with our formal asymptotic analysis. The function $G_{\alpha,0}$, see (6.8), is not defined if $\psi'(x_\alpha(t)) = 0$, which is the case if $x_\alpha \in \{x_*, x^*\}$. Hence we can not derive anything for the possible change of $\lambda$ in these cases. The experiments in Section 5.3 suggest, that the phase fraction $\lambda$ changes in time when the position $x_L(t)$ or $x_R(t)$ reaches the position of local extrema in $\psi'$, i.e. $x_*$ or $x^*$. This would then give the same behaviour as described for the discrete many particle system in Section 2.3. We expect the two limiting behaviours to coincide. In a new work by Herrmann, Niethammer and Velázquez, [19], the authors are able to derive formally the evolutionary behaviour for the phase fraction $\lambda$ and the positions $x_L$ and $x_R$. They work on an unbounded domain, which can be seen as a technical detail, and we expect that their results are valid for our case as well. They also identify other limit behaviour for different scalings, i.e. different relations of the sequences $\tau_k \to 0$ and $\nu_k \to 0$.

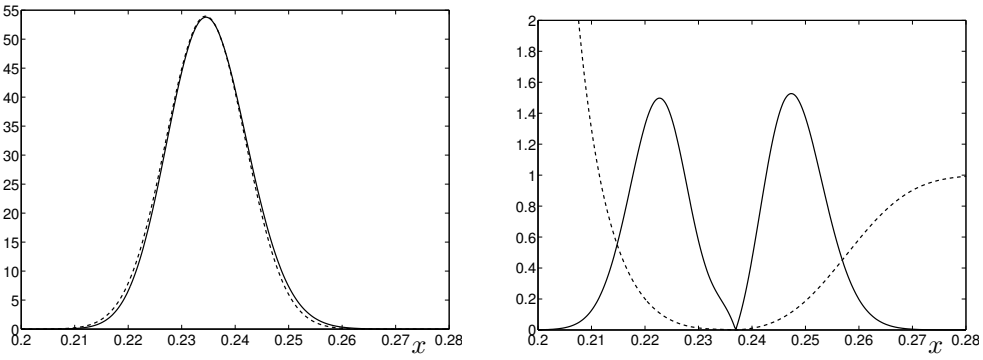



Figure 6.4: First order approximation with formal asymptotics, left: solution of experiment in Figure 5.7 at time $t_2$ (solid) and its first order approximation (dashed), right: absolute error of approximation (solid) and relative error (dashed).

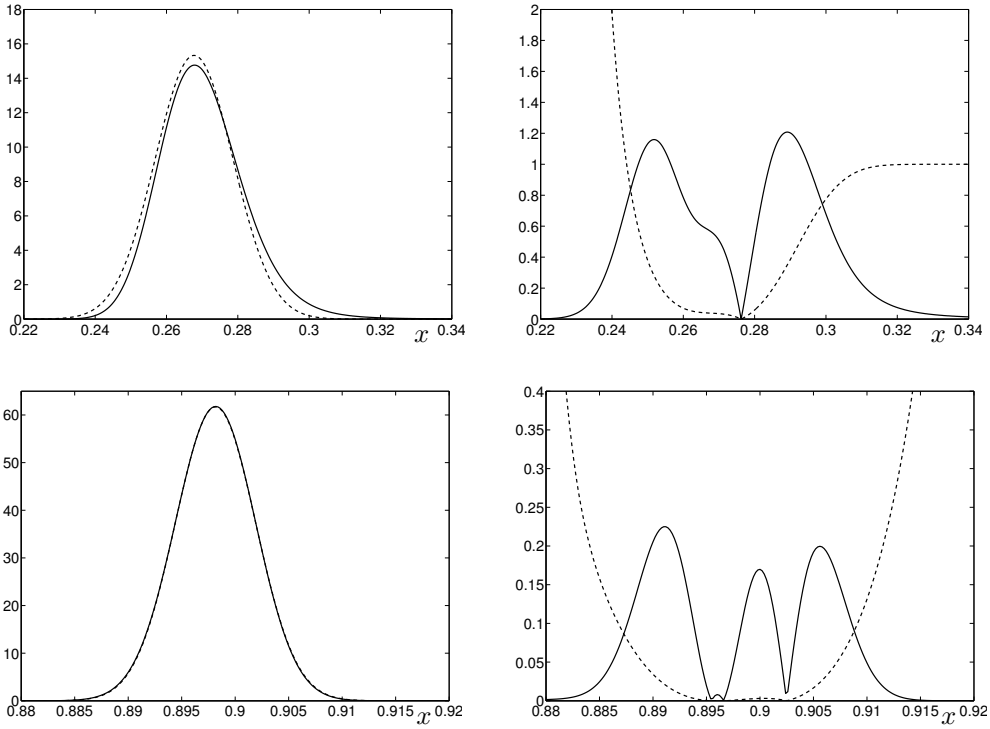



Figure 6.5: First order approximation with formal asymptotics at the transition time, left (top and bottom): solution of experiment in Figure 5.7 at time $t_4$ (solid) and its first order approximation (dashed), right (top and bottom): absolute error of approximation (solid) and relative error (dashed), the value $\lambda(t_4)$ is guessed from the masses of left and right side of the solution.

# 7 Conclusion and Outlook

In order to conclude the results, this work must be looked at from two different points of view, of which the first is the mathematical one. Here we have investigated a nonlinear PDE numerically and analytically. From a physical point of view two different models, which describe the charging of a lithium ion battery were discussed.

We showed existence and uniqueness of solutions to a nonlinear Fokker–Planck equation with the help of interpolation spaces and semigroup techniques. For the global existence in time we exploited the strong connection to an energy functional as used in the context of Wasserstein gradient flows. The necessary and sufficient conditions for global existence of solutions is congruent with what one would expect from a physical point of view. We also present a convergence result for the numerical solutions which gives confidence into the numerical experiments. On the one hand we do not derive any convergence rates. On the other hand we do not rely on high regularity of the exact solution. The strength of the numerical convergence result is that it can deal with a nonlinear term that depends on the solution as a higher order function. Hence, we formulated this result in a general form making it applicable to other semilinear parabolic problems.

The interesting behaviour of convergence to Dirac measures was documented numerically and we also gave evidence by the method of formal matched asymptotics. The solution tends to develop on a fast time-scale typical profiles or patterns, which then change on a comparable slow time scale. This phenomenon is similar to the *metastable patterns* in solutions to the Allen-Cahn equation described by Carr and Pego in [5] or by Chen in [6]. For the Allen-Cahn equation one exploits, that it is an $L^2(\Omega)$ gradient flow of some energy and then shows, that there is a strongly attracting manifold of solutions that have the above mentioned patterns. It would be interesting to see if a similar approach is possible when exploiting a Wasserstein gradient-flow structure in the nonlinear Fokker-Planck equation at hand.

From the physical point of view, we analysed and compared a *discrete many particle model* and a *continuous many particle model* for the loading of a lithium-ion battery. The *discrete many particle model* described in Section 2.3 shows, that the two voltage plateaus while charging and discharging as well as their different heights stem from the interaction of the many nanosized lithium storage particles in the battery's cathode. We conclude that most of the loading states of the battery can only be realised with two distinct groups of particles having high and low relative loading states. In quasi-static loading this explains the almost horizontal voltage plateaus while (dis)charging. The investigated model, however, cannot satisfactorily explain how one can influence the difference in heights of the voltage plateaus while charging and furthermore it seems hard to expand this model to a non-quasi-static setting. This serves as a motivation for

the *continuous many particle model*.

Starting from Chapter 3 we focussed on the *continuous many particle model*, which arises from the *discrete many particle model* in view of a large number of particles. The model gives a nonlinear PDE and we showed the existence and uniqueness for solutions for all physically meaningful data. The PDE reflects the formation of two different groups of particles with high and low relative loading state like the *discrete many particle model*. Additionally, we documented the influence of the appearing parameters on the shape of the charging and discharging curves.

At the end of Chapter 2 we find that the *discrete many particle model* can be modified into a stochastic differential equation. We strongly believe that this can result directly into the *continuous many particle model*. Moreover it would be interesting to extend the *continuous many particle model* and drop some simplifying assumptions to capture more of the underlying processes. This can be done, for example, via the introduction of more dimensions in the property space, in order to reflect the intercalation processes inside the particles. The models presented here are reduced to the description of the cathode of a battery only and so, a next interesting step could be to embed the modelling of the cathode into a larger battery model.

# Appendix

We use this place to provide some frequently used estimates in this work. Even though we state most without proof, they are collected here for the convenience of the reader.

## A Estimates

The following lemma presents a Gagliardo Nirenberg estimate.

**Lemma A.1.** *(i) Let $\alpha$, $\beta$ be integers, satisfying $0 \leq \alpha < \beta$ and let $1 \leq q, r \leq \infty$, $0 \leq p < \infty$. For the case $q$ or $r$ having the value $\infty$, we define formally $\frac{1}{\infty} = 0$. Then we define $\theta$ as*

$$\theta := \frac{\frac{1}{p} - \frac{1}{q} - \alpha}{\frac{1}{r} - \frac{1}{q} - \beta}.$$

*If $\theta \in [\frac{\alpha}{\beta}, 1]$, then there exist constants $c_0, c_1 \geq 0$ such that for all $\varphi \in \mathrm{H}^{\beta,r}(\Omega) \cap \mathrm{L}^q(\Omega)$ there holds*

$$\|\frac{\partial^\alpha}{\partial x^\alpha}\varphi\|_{\mathrm{L}^p} \leq c_0 \|\frac{\partial^\beta}{\partial x^\beta}\varphi\|_{\mathrm{L}^r}^\theta \|\varphi\|_{\mathrm{L}^q}^{1-\theta} + c_1 \|\varphi\|_{\mathrm{L}^q}. \tag{1}$$

*(ii) For all $0 < q < \infty$ and $0 < r \leq \infty$ there exists $c > 0$, such that for all $\varphi \in \mathrm{H}^1(\Omega) \cap L^q(\Omega) \cap L^r(\Omega)$ there holds*

$$\|\varphi\|_{\mathrm{C}([0,1])} \leq (\frac{q}{2} + 1)^\theta \|\varphi_x\|_{L^2}^\theta \|\varphi\|_{L^q}^{1-\theta} + \|\varphi\|_{L^r}, \tag{2}$$

$$\textit{and} \qquad \|\varphi\|_{\mathrm{C}([0,1])} \leq (\frac{q}{2} + 1)^\theta (\|\varphi\|_{L^2} + \|\varphi_x\|_{L^2})^\theta \|\varphi\|_{L^q}^{1-\theta}, \quad \textit{with} \quad \theta = \frac{2}{q+2}. \tag{3}$$

*Proof.* (i)This statement is taken from [43, Theorem 1.3.4].
(ii) We know that $\mathrm{H}^1(\Omega) \subset \mathrm{C}(\bar{\Omega})$, such that we can define $x_*, x^* \in [0,1]$ as

$$|\psi(x_*)| \leq |\psi(x)| \leq |\psi(x^*)| \qquad \forall x \in [0,1].$$

Then for all $\beta > 1$ there holds

$$\begin{aligned} \|\psi\|_{L^\infty}^\beta = |\psi(x^*)|^\beta &\leq \Big| \int_{x_*}^{x^*} \left(|\psi|^\beta\right)_x \,\mathrm{d}x \Big| + |\psi(x_*)|^\beta \\ &\leq \beta \int_0^1 |\psi|^{\beta-1} |\psi_x| + |\psi(x_*)|^\beta \,\mathrm{d}x \leq \beta \|\psi\|_{L^{2\beta-2}}^{\beta-1} \|\psi_x\|_{L^2} + |\psi(x_*)|^\beta. \end{aligned} \tag{4}$$

Applying the bound $|\psi(x_*)| \le \|\psi\|_{L^r}$ and setting $\beta = \frac{q}{2} + 1 > 1$, this proves (2). On the other hand, keeping the choice of $\beta$ we can proceed from (4) with

$$\begin{aligned}\|\psi\|_{L^\infty}^\beta &\le \beta\|\psi\|_{L^{2\beta-2}}^{\beta-1}\|\psi_x\|_{L^2} + |\psi(x_*)|^{\beta-1}|\psi(x_*)| \\ &\le \beta\|\psi\|_{L^{2\beta-2}}^{\beta-1}\|\psi_x\|_{L^2} + \|\psi(x)\|_{L^{2\beta-2}}^{\beta-1}\|\psi(x)\|_{L^2} \le \beta\|\psi\|_{L^{2\beta-2}}^{\beta-1}(\|\psi\|_{L^2} + \|\psi_x\|_{L^2}).\end{aligned}$$

This then proves (3). □

We also provide a Gagliardo-Nirenberg type estimate involving norms in $L\log L(\Omega)$. The proof consists of a modification of [3, p. 1199].

**Lemma A.2.** *Let $\Omega \subset \mathbb{R}$ be a bounded interval. There exists $C > 0$ with the property that for all $\varepsilon > 0$ one can find $C_\varepsilon > 0$, such that*

$$\|w\|_{\mathrm{L}^\infty}^3 \le \varepsilon\|\partial_x w\|_{\mathrm{L}^2}^2 \cdot \|w\ln|w|\|_{\mathrm{L}^1} + C_\varepsilon + C\|w\|_{\mathrm{L}^1}^3 \tag{5}$$

*is valid for all $w \in \mathrm{H}^1(\Omega)$.*

*Proof.* Following the reasoning in [3], we first invoke the Gagliardo-Nirenberg inequality (2) to find $c_1 > 0$ such that

$$\|v\|_{\mathrm{L}^\infty}^3 \le c_1\|\partial_x v\|_{\mathrm{L}^2}^2 \cdot \|v\|_{\mathrm{L}^1} + c_1\|v\|_{\mathrm{L}^1}^3 \qquad \text{for all } v \in \mathrm{H}^1(\Omega). \tag{6}$$

We now choose $N > 1$ large fulfilling $\frac{8c_1}{\ln N} \le \varepsilon$ and introduce $\chi \in \mathrm{W}^{1,\infty}_{loc}(\mathbb{R})$ by defining $\chi(s) := 0$ for $s \in [-N, N]$, $\chi(s) := |s|$ for $|s| \ge 2N$ and $\chi(s) := 2(|s| - N)$ for $N < |s| < 2N$. Then given $w \in \mathrm{H}^1(\Omega)$, we evidently have

$$\|w - \chi(w)\|_{\mathrm{L}^\infty} \le 2N$$

and furthermore

$$\|\chi(w)\|_{\mathrm{L}^1} \le \int_{\{|w|>N\}} |w|\,\mathrm{d}x \le \frac{1}{\ln N} \cdot \|w\ln|w|\|_{\mathrm{L}^1}.$$

Since $(1+\xi)^3 \le 2\cdot(1+\xi^3)$ for $\xi \ge 0$, (6) furthermore yields

$$\begin{aligned}\|w\|_{\mathrm{L}^\infty}^3 &\le 2\|\chi(w)\|_{\mathrm{L}^\infty}^3 + 2\|w - \chi(w)\|_{\mathrm{L}^\infty}^3 \\ &\le 2c_1\|\partial_x\chi(w)\|_{\mathrm{L}^2}^2 \cdot \|\chi(w)\|_{\mathrm{L}^1} + 2c_1\|\chi(w)\|_{\mathrm{L}^1}^3 + 2^4N^3 \\ &\le \tfrac{8c_1}{\ln N} \cdot \|\partial_x w\|_{\mathrm{L}^2}^2 \cdot \|w\ln|w|\|_{\mathrm{L}^1} + 2c_1\|w\|_{\mathrm{L}^1}^3 + 3^4N^3,\end{aligned}$$

because $\|\chi'\|_{\mathrm{L}^\infty(\mathbb{R})} = 2$ and $|\chi(s)| \le |s|$ for all $s \in \mathbb{R}$. In view of our definition of $N$, this proves (5) with $C := 2c_1$ and $C_\varepsilon := 2^4N^3$. □

Some frequently used estimates in this work are for the semigroup generated by sectorial operators. For this we first define the notion of a sectorial operator. Let $X$ be a Banach space. We consider an operator which is defined on its domain $D(A) \subset X$ and maps into $X$. We say that $A$ is sectorial, if for a real number $\omega$ and an angle $\theta \in ]\pi/2, \pi[$

its resolvent $R(\lambda, A) := (A - \lambda I)^{-1}$ exists for all

$$\lambda \in S_{\theta,\omega} := \{\phi \in \mathbb{C} : \lambda \neq \omega, |\arg(\phi - \omega)| \leq \theta\} ,$$

and furthermore there exists $m > 0$, such that the resolvent is a linear and compact operator and bounded in the form

$$\|R(\lambda, -)v\|_X \leq \frac{m}{|\lambda - \omega|}\|v\|_X \qquad \text{for all } v \in V, \lambda \in S_{\theta,\omega} .$$

For sectorial operators one can define for all $t > 0$ the the operator $\mathrm{e}^{-At} : X \to X$, see the books by Pazi [33] and Lunardi [28]. The following stems from those books.

**Lemma A.3.** *For a sectorial operator $A$ there is for all $\gamma > 0$ a constant $c$ such that for all $v \in X$ there holds*

$$\|A^\gamma \mathrm{e}^{-At}v\|_X \leq ct^{-\gamma}\mathrm{e}^{\omega t}\|v\|_X .$$

*The constant $c$ depends on $\theta$, $m$, $\theta$ and $\gamma$.*

*Proof.* Proof can be found in [33] or [28]. □

# Bibliography


[1] Steffen Arnrich, Alexander Mielke, Mark Peletier, Giuseppe Savaré, and Marco Veneroni, *Passing to the limit in a wasserstein gradient flow: from diffusion to reaction*, Calculus of Variations and Partial Differential Equations (2011), 1–36, to appear.

[2] Nikolai Yu. Bakaev, *Analysis of fully discrete approximations to parabolic problems with rough or distribution-valued initial data*, Numerische Mathematik **92** (2002), no. 4, 621–651.

[3] P. Biler, W. Hebisch, and T. Nadzieja, *The Debye system: Existence and large time behavior of solution*, Nonlinear Analysis, TMA **23** (1994), no. 9, 1189–1209.

[4] James H. Bramble, Joseph E. Pasciak, and Olaf Steinbach, *On the stability of the $L^2$ projection in $H^1(\Omega)$*, Math. Comput. **71** (2002), 147–156.

[5] J. Carr and R. L. Pego, *Metastable patterns in solutions of $u_t = \varepsilon^2 u_{xx} - f(u)$*, Communications on Pure and Applied Mathematics **42** (1989), no. 5, 523–576.

[6] Xinfu Chen, *Generation, propagation, and annihilation of metastable patterns*, Journal of Differential Equations **206** (2004), no. 2, 399–437.

[7] Philippe G. Ciarlet, *Finite element method for elliptic problems*, Classics in Applied Mathematics, no. 40, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2002.

[8] C. Delmas, M. Maccario, L. Croguennec, F. Le Cras, and F. Weill, *Lithium deintercalation in $LiFePO_4$ nanoparticles via a domino-cascade model*, Nat Mater **7** (2008), no. 8, 665–671.

[9] W. Dreyer, M. Gabercek, C. Guhlke, R. Huth, and J. Jamnik, *Phase transition in a rechargeable lithium battery*, European Journal of Applied Mathematics **22** (2011), no. 03, 267–290.

[10] W. Dreyer, I. Müller, and P. Strehlow, *A study of equilibria of interconnected balloons*, The Quarterly Journal of Mechanics and Applied Mathematics **35** (1982), no. 3, 419–440.

[11] Wolfgang Dreyer, Clemens Guhlke, and Michael Herrmann, *Hysteresis and phase transition in many-particle storage systems*, Continuum Mechanics and Thermodynamics **23** (2011), 211–231.

[12] Wolfgang Dreyer, Clemens Guhlke, and Robert Huth, *The behavior of a many-particle electrode in a lithium-ion battery*, Physica D: Nonlinear Phenomena **240** (2011), 1008–1019.

[13] Wolfgang Dreyer, Robert Huth, Alexander Mielke, Joachim Rehberg, and Michael Winkler, *Blow-up versus boundedness in a nonlocal and nonlinear fokker-planck equation*, WIAS Preprint 1604, 2011.

[14] Wolfgang Dreyer, Janko Jamnik, Clemens Guhlke, Robert Huth, Joze Moskon, and Miran Gaberscek, *The thermodynamic origin of hysteresis in insertion batteries*, Nat Mater **9** (2010), no. 5, 448–453.

[15] M. J. Englefield, *Exact solutions of a fokker-planck equation*, Journal of Statistical Physics **52** (1988), no. 1, 369–381.

[16] Lawrence C. Evans, *Partial Differential Equations (Graduate Studies in Mathematics, V. 19) GSM/19*, American Mathematical Society, June 1998.

[17] Kramers H.A., *Brownian motion in a field of force and the diffusion model of chemical reactions*, Physica **7** (1940), no. 4, 284–304.

[18] E. Hairer and G. Wanner, *Solving ordinary differential equations: Stiff and differential-algebraic problems*, vol. II, Springer series in computational mathematics, no. 14, Springer-Verlag, 1993.

[19] M. Herrmann, B. Niethammer, and J. J. L. Velázquez, *Kramers and non-Kramers Phase Transitions in Many-Particle Systems with Dynamical Constraint*, ArXiv e-prints (2011), 1–33, arXiv:1110.3518v1.

[20] J.K. Hunter and B. Nachtergaele, *Applied analysis*, World Scientific, 2001.

[21] Claes Johnson, Stig Larsson, Vidar Thomée, and Lars B. Wahlbin, *Error estimates for spatially discrete approximations of semilinear parabolic equations with nonsmooth initial data*, Mathematics of Computation **49** (1987), no. 180, 331–357.

[22] Richard Jordan, David Kinderlehrer, and Felix Otto, *The variational formulation of the Fokker-Planck equation*, SIAM Journal on Mathematical Analysis **29** (1998), no. 1, 1–17.

[23] J.L. Kelley, *General topology*, Graduate texts in mathematics, Springer-Verlag, 1975.

[24] A. Kolmogoroff, *Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung*, Mathematische Annalen **104** (1931), 415–458, Berlin.

[25] P.A. Lagerstrom, *Matched asymptotic expansions - ideas and techniques*, Applied Mathematical Sciences, vol. 76, Springer-Verlag New York, September 1988.

[26] S. Larsson and V. Thomée, *Partial differential equations with numerical methods*, Texts in applied mathematics, Springer, 2003.

[27] Christian Lubich and Alexander Ostermann, *Runge-kutta time discretization of reaction-diffusion and navier-stokes equations: nonsmooth-data error estimates and applications to long-time behaviour*, Applied Numerical Mathematics **22** (1996), no. 1-3, 279–292.

[28] Alessandra Lunardi, *Analytic semigroups and optimal regularity in parabolic problems*, Progress in Nonlinear Differential Equations and their Applications, vol. 36, Birkhäuser, 1995.

[29] Thomas Maxisch and Gerband Ceder, *Elastic properties of olivine* $Li_xFePO_4$ *from first principles*, Phys. Rev. B **73** (2006), 174112.

[30] Alexander Mielke and Lev Truskinovsky, *From discrete visco-elasticity to continuum rate-independent plasticity: Rigorous results*, Archive for Rational Mechanics and Analysis (To appear 2011), 1–43.

[31] Ingo Müller, *Grundzüge der Thermodynamik*, 3. ed., Springer, 2001, ISBN 978-3-540-42210-5.

[32] A. K. Padhi, K. S. Nanjundaswamy, and J. B. Goodenough, *Phospho-olivines as positive-electrode materials for rechargeable lithium batteries*, Journal of The Electrochemical Society **144** (1997), no. 4, 1188–1194.

[33] A. Pazy, *Semigroups of linear operators and applications to partial differential equations*, Applied Mathematical Sciences, vol. 44, Springer-Verlag, New York, 1983.

[34] Mark A. Peletier, Giuseppe Savaré, and Marco Veneroni, *From diffusion to reaction via* $\Gamma$*-convergence*, SIAM J. Math. Anal **42** (2010), no. 4, 1805–1825.

[35] G. Puglisi and L. Truskinovsky, *Rate independent hysteresis in a bi-stable chain*, Journal of Mechanics and Physics of Solids **50** (2002), 165–187.

[36] Nathalie Ravet, Simon Besner, Martin Simoneau, Alain Vallée, Michel Armand, and Jean-François Magnan, *Electrode materials with high surface conductivity*, CA Patent No. EP1049182, 2000.

[37] M. Safari and C. Delacourt, *Mathematical modeling of lithium iron phosphate electrode: Galvanostatic charge/discharge and path dependence*, Journal of The Electrochemical Society **158** (2011), no. 2, A63–A73.

[38] Venkat Srinivasan and John Newman, *Discharge model for the lithium iron-phosphate electrode*, Journal of The Electrochemical Society **151** (2004), no. 10, A1517–A1529.

[39] ———, *Existence of path-dependence in the* $LiFePO_4$ *electrode*, Electrochemical and Solid-State Letters **9** (2006), no. 3, A110–A114.

[40] A. Visintin, *Strong convergence results related to strict convexity*, Communications in Partial Differential Equations **9** (1984), no. 5, 439–466.

[41] Marnix Wagemaker, Wouter J. H. Borghols, and Fokko M. Mulder, *Large impact of particle size on insertion reactions. a case for anatase $Li_x TiO_2$*, Journal of the American Chemical Society **129** (2007), no. 14, 4323–4327.

[42] Wen-Chyuan Yueh, *Eigenvalues of several tridiagonal matrices*, Appl Math ENotes **5** (2005), 66–74.

[43] S. Zheng, *Nonlinear evolution equations*, Chapman & Hall/CRC monographs and surveys in pure and applied mathematics, Chapman & Hall/CRC Press, 2004.

# List of Figures

# Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 11.01.2012 Robert Huth